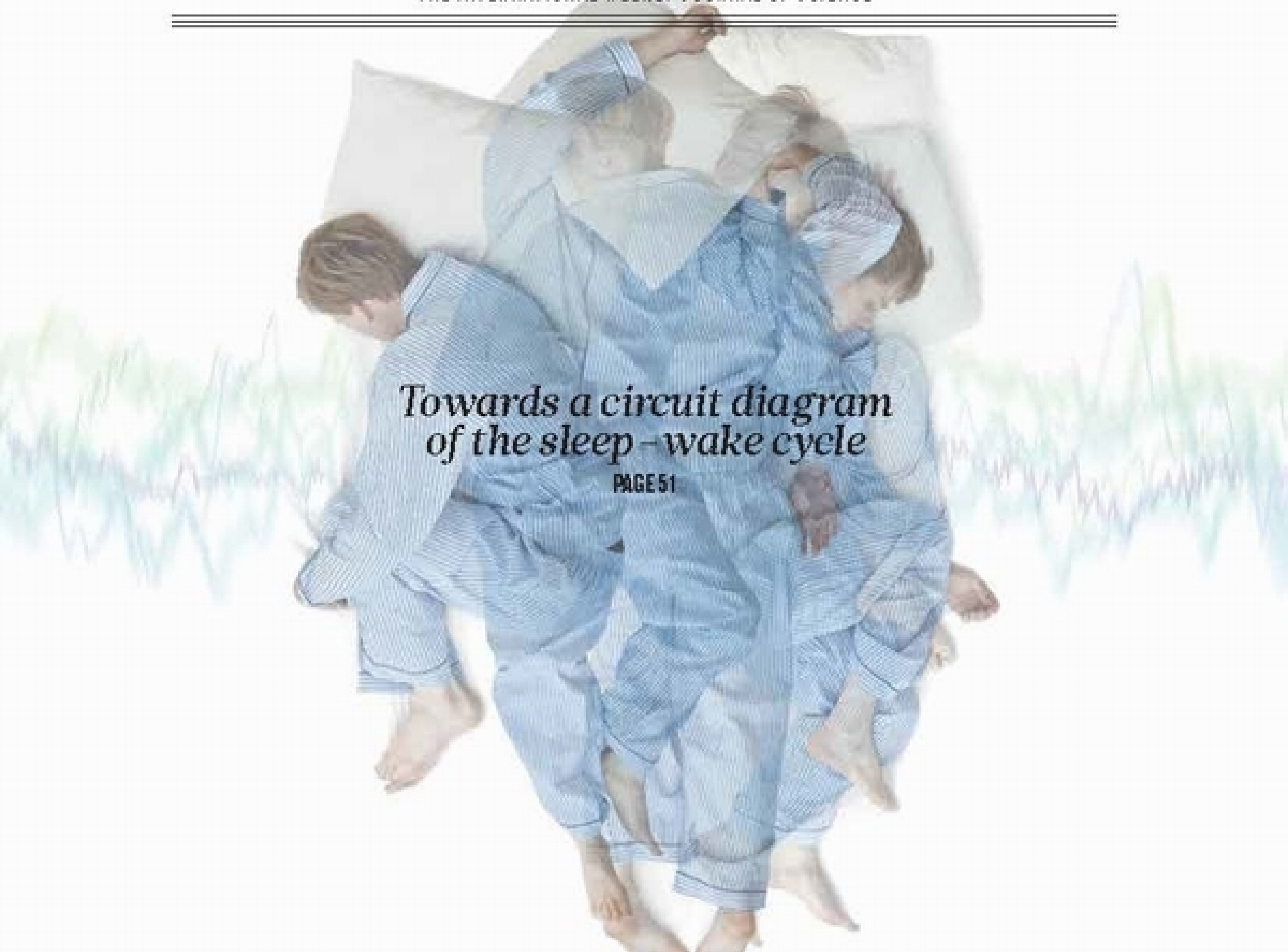
nature

THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE



Towards a circuit diagram of the sleep–wake cycle

PAGE 51

Taking control of sleep

READING

BOOK REVIEWS SPECIAL

Fashions in physics, Nazis on drugs and cybernetics

PAGE 32

ENERGY

CLEANING UP OUR ACT

Agree priorities, baselines for renewables R&D

PAGES 27 & 30

CLIMATE

CH_4 BUDGET REBALANCED

Methane emissions from fossil fuels revised upwards

PAGES 46 & 88

NATURE.COM/NATURE

6 October 2016 £10

Vol. 538, No. 7623

9 770028 083095 40

THIS WEEK

EDITORIALS

LIFESPAN The secret to a not-so-long but happy life **p.6**

WORLD VIEW Kick corporate culture out of universities **p.7**

BIOSHIELD Skin microbe protects toads from fungus **p.9**

Stand up for UK research freedom

A proposed higher-education and research bill would demolish the agreements that protect British universities from political interference. It must be opposed.

As publicly funded employees, British academic scientists are fortunate. They surfed the high-spending wave of the Labour government years, starting in the late 1990s. When the 2008 financial crisis hit, they were protected from the deep public-sector budget cuts that followed. Public libraries closed. Some of the poorest people lost welfare benefits, and university students faced trebling tuition fees. But for established researchers it was, on balance, business as usual. Now, that relative stability is set to change.

A draft law, the Higher Education and Research Bill, is making its way through the House of Commons. The bill amounts to the biggest shake-up in the sector for more than a generation. It is designed, among other things, to make it easier for private companies to set up universities, and to enable more researchers to commercialize their work. If it passes, existing funding bodies will close and replacements will be created. But in the process of change, the bill rips up an 800-year-old settlement between the nation's scholars and the state. It opens the door to unacceptable political interference. It must be resisted.

At the moment, scientists have a right in law to choose what to work on without unwarranted steering or instruction from government. This protection for the integrity of scholarship is enshrined in a centuries-old legal instrument called a royal charter. First used before the United Kingdom's parliamentary system was established, royal charters keep public bodies (including the BBC) at arm's length from meddling ministers, and so shield their activities from the prevailing — and changeable — political winds. Many scientists may not know it, but the royal charters of their universities help public funds for research and teaching to come with few strings attached.

The University of Cambridge received its royal charter in 1231, and dozens of other universities have been granted them since. Royal charters also govern each of the seven discipline-based research councils.

The UK government's proposed law would change that. The bill would dissolve the seven individual research funding councils; the body that would replace them, called UK Research and Innovation, would have no royal charter.

The bill also proposes to override the royal charters of universities. This would happen with the establishment of another governmental body, the Office for Students. This would regulate the expected flood of new private universities, as well as existing publicly funded ones. So even for those universities that have a royal charter, the creation of the Office for Students would effectively make that document worthless.

Why does this matter? As the draft legislation makes clear, ministers would then be able to suggest courses for universities to teach. Furthermore, the government would give itself the direct right to create and dissolve whole areas of research funding. At present, the risk to the autonomy of science and research is theoretical — but the implications for academic freedom are troubling.

So far, there has been little sign of resistance from members of Parliament (MPs). The opposition Labour Party is engulfed in a divisive civil war and has not been able to focus properly on the bill, despite the best efforts of its science and higher education team. The government, meanwhile, has convinced its own rebellious MPs to support the bill.

Organizations representing scientists, along with pressure groups such as the Campaign for Science and Engineering in London, have largely maintained public silence. That is understandable to an extent, because they are used to having a positive relationship with ministers and are more experienced at advocating for their causes in private meetings.

"A government that is determined to have its way needs to be confronted in public."

But a government that is determined to have its way needs to be dealt with differently. It needs to be confronted in public.

That could happen as soon as this month, when the bill will be discussed in the House of Lords. Several research and higher-education leaders who now sit in the Lords plan to give the bill more forensic scrutiny than it has received in the Commons. However, without wider and more vocal support from the science community, their efforts will be no more than an inconvenient blip in the bill's journey into law.

Make no mistake. Britain's first all-Conservative government in 20 years sees science and higher education as vestiges of the big state. If its proposals become law, the government will upend globally accepted norms that protect independence and self-determination in science and higher education. If scientists and their representative organizations don't want that to happen, they need to speak up — and do it now. ■

A good prize

Nobel awards week shows the value of a strong brand identity.

As befits someone who made his fortune from dynamite, Alfred Nobel was worried about a premature death. The will that set up prizes in his name is most well known for his much discussed — if vague — intention that the awards should recognize work with a benefit for humanity. Less well known is that the will concludes with an instruction from Alfred for a doctor to open his veins, allow him to bleed out, and then, unusually for the time, to burn his remains in a new-fangled crematorium. This was a man determined to avoid being buried alive. (Given his fear of being wrongly diagnosed as deceased, it must have been a shock for him to read his

own obituary, published in error on the death of his brother almost a decade before his own death.)

Nobel prize week is a time when some showbiz glamour is sprinkled on the world of science and research. For a few days each year, the names and photographs of scientists are presented to the public, alongside — sometimes surprisingly detailed — descriptions of their discoveries and the benefits they provide. Already this week, analyses of the cellular mechanism of autophagy (or how cells digest and recycle their components) and of exotic states of matter that may pave the way for quantum computers have been laid out for public consumption (see page 18).

In a world of increasing competition for eyeballs, attention and web clicks, it's worth remembering that the Nobel prizes are a global, regular and almost-universally admired advertisement for the career that many of *Nature*'s readers dedicate their lives to — and frequently lament that the wider public does not appreciate.

That's not to say that the Nobel prizes are immune from criticism. Do Alfred's original categories truly reflect the span of modern science? And why limit the number of prizewinners to three? Readers with a taste for Counter-Reformation baroque Flemish art can enjoy a lengthy defence of the three-prize limit that was published in the journal *Cell* last month (J. L. Goldstein *Cell* **167,** 5–8; 2016), in which the author eagerly draws on the triptych paintings of Peter Paul Rubens (and later Francis Bacon) for inspiration. More tangibly perhaps, the untimely death of physicist Deborah Jin has refocused debate on the extent to which the annual decisions of the Nobel prize committee should be swayed by whether deserving candidates will be alive to receive an award in future years. (The rules laid out in Alfred's will state that prizes cannot be awarded posthumously.)

The proliferation of academic prizes in recent years — some of which are much more lucrative than the Nobels — has increased the pressure on the Nobel Foundation to move with the times. It's what corporate brand consultants call a clash between identity — what an organization chooses to do — and reputation, or how that action sits with what people on the outside think it should do.

"Nobel prize week is a time when some showbiz glamour is sprinkled on the world of science and research."

But as one Nobel official puts it: "I don't think the reputation of the Nobel prize was built by people caring about the reputation of the prize." And, for good measure, he adds: "It is not necessarily a remit to go out and find out what the world thinks of the Nobel prize and try and adjust our behaviour because of that … It is interesting to know what the world thinks of the Nobel prize, but should that change our behaviour?"

There is a motto at the Nobel Foundation: a good prize one year will be a better one the next. So far, it is difficult to argue with the benefit. ■

Dance with death

The search for eternal life could be scuppered by the limits of the human body.

Why do animals grow old and die at characteristic ages? Even if maintained in peak condition and not eaten by your cat, your hamster is unlikely to make it much past its second birthday. And your cat might live for ten times that. Yet neither cat nor hamster will ever match the average healthy human for longevity.

A study published online in *Nature* this week uses demographic data to reveal a lifespan that human beings cannot exceed, simply by virtue of being human (see X. Dong *et al. Nature* http://dx.doi.org/10.1038/nature19793; 2016). It's like running, as an accompanying News and Views article points out (see S. J. Olshansky *Nature* http://dx.doi.org/10.1038/nature19793; 2016). Elite athletes might shave a few milliseconds off the world record for the 100-metre sprint, but they'll never run the same distance in, say, five seconds, or two. Human beings are simply not made that way. The same is true for longevity. The consequences of myriad factors related to our genetics, metabolism, reproduction and development, all shaped over millions of years of evolution, means that few humans will make it past their 120th birthdays. The name of Jeanne Calment, who died in 1997 at the age of 122, is likely to remain as long in the memory in the Methuselah stakes as that of Usain Bolt on the Olympic track.

Maximum lifespan is a bald measure of years accumulated. It is not the same as life expectancy, which is an actuarial measure of how long one is expected to live from birth, or indeed from any given age. Life expectancy at birth has increased in most countries over the past century, not because people have longer lifespans, but mainly because infectious disease does not kill as many infants as it once did. Factors such as poverty and warfare conspire to decrease life expectancy. Although life expectancy at birth has risen steadily for both men and women in France since 1900, for example, there are dramatic and poignant drops that coincide with the two world wars.

In Britain in the early twentieth century, many children still died from infectious diseases, and men would die shortly after retiring from physically demanding jobs. The National Health Service was the political response. It has become, in some ways, the victim of its own success. People live longer than they did even a few decades ago, and die (eventually) of different (and more expensive) complaints. As any beginning medical student is soon taught, gerontology is far from a dying discipline. So if we owe our increases in life expectancy to better public health, nutrition, sanitation and vaccination, is it not fair to ask whether more-effective treatments for diseases such as cancer, Parkinson's disease and Alzheimer's might also yield dividends in maximum lifespan? Will 120th birthday parties become routine, outmatched by a small yet increasing number of sesquicentenarians? The demographic data say no. People are living longer, and the population as a whole is greying, but the rate of increase in the number of centenarians is slowing, and might even have peaked.

Could it be possible, in some science-fictional future, to break free from the bonds of human life expectancy and increase lifespan indefinitely? An unquenchable desire for eternal life has preoccupied humanity from the earliest times, as attested by the earliest passages of the Bible, the Gilgamesh epic and many other stories from our past. Perhaps the chilliest evocation of mortality comes in Bede's seventh-century *Ecclesiastical History of the English People*, in which a chieftain remarks that the 'few moments of comfort' offered by human life are as the brief flight of a sparrow through a warm and lighted mead hall, in through one door, and out through the other, back into a dark, storm-tossed and demon-haunted night of which we know nothing. No wonder we'd all like a little more light. Technological solutions might one day transcend the limitations of the human body, but transcend them they must — mere extension is already yielding diminishing returns.

The risks of transcendence are twofold. First, it might be that to extend our lives beyond our normal span, we must somehow become other than human. After all, what would a 50-year-old hamster be like? The unintended consequences of immortality are graphically and grimly illustrated in Aldous Huxley's 1939 novel *After Many A Summer,* in which people fed on a life-extending diet of carp intestines live for centuries — at the cost of turning into witless apes. Second, there is a risk that life wouldn't really be that much longer — it would only feel like it. ■

Corporate culture has no place in academia

'Academic capitalism' contributed to the mishandling of the Macchiarini case by officials at the Karolinska Institute in Sweden, argues **Olof Hallonsten**.

The eyes of the world are on the Karolinska Institute in Stockholm this week for the announcement of the Nobel prize in medicine, but an ugly medical scandal lurks in the background. The case of Paolo Macchiarini involved the deaths of multiple patients and several instances of research fraud, and has exposed the misdeeds of a single professor. But it also demonstrates the risks of academic capitalism: a global trend that turns universities into businesses. In this respect, the story has wider lessons for us all.

As academic capitalism spreads, universities abandon traditional meritocratic and collegial governance to hunt money, prestige and a stronger brand. Here in Sweden, this shift has been especially profound: since the 1980s, the university system has been deregulated, and its core principles gradually replaced by management practices from the corporate world. Government research policy over the past decade has further pushed universities to centralize their strategic management and increase their international visibility. Major strategic funding programmes included one to recruit international star scientists.

An investigation into the Macchiarini scandal, led by a former president of the Supreme Administrative Court of Sweden, Sten Heckscher, delivered its report last month, and puts some blame on this "new orientation of research policy". There is now an elevated risk that fraud is not properly detected and that ethically doubtful research is allowed to continue, notes the report, because new policy incentives cloud the judgement of academic leaders.

The Heckscher investigation shows how officials at the Karolinska Institute (KI) contributed to the scandal. In their efforts to recruit Macchiarini in 2010, and in their handling of the renewal of his contract and the allegations against him in 2011–15, university leaders short-circuited regulations and established practices. They failed to have Macchiarini's research properly peer-reviewed, and ignored both allegations of research fraud and the results of external investigations.

From the outside, it seems that KI officials were tempted by the prospect that Macchiarini would revolutionize regenerative medicine and thus bring great prestige and worldwide acclaim. Already placed highest among Swedish universities on global ranking lists, the KI no doubt saw a chance to distinguish itself further and attract more funding and prestige, in an endless hunt for greater acclaim.

Yet this conduct goes against fundamental values of academia — the careful scrutiny of all claims, and of the research (and teaching) portfolios of those making such claims. This core principle in the self-organization of the academic system (studied by sociologists Robert Merton and Pierre Bourdieu, among others) is intended to guarantee that science progresses and delivers knowledge and technology to society that is as accurate as possible and not gained unethically.

Academic capitalism runs counter to these ideals, subsuming achievement in research and teaching to attainment of economic goals and quantitatively oriented (and shallow) performance assessments and rankings. Academic self-regulation and vocational autonomy are replaced with external control by audit and management. The individual's struggle for recognition in science is colonized by university managers, who use the achievements of scientists and students to accumulate capital (economic, symbolic and cultural, in Bourdieu's terms), and thus increase the visibility of their university.

As Heckscher's investigation shows, the intervention of the KI's then rector, Anders Hamsten, in the renewal of Macchiarini's contract with the university in 2015 led to the arbitrary acquittal of Macchiarini from accusations of scientific fraud in the same year. Thus acts an academic leader who has abandoned sound academic practice in favour of maximizing the prestige and finances of his university. An academic leader remaining true to the classic ideals, and embedded in a sound academic culture and research policy, would have made the obvious choice to investigate the fraud allegations thoroughly at an early stage, looking beyond the count of publications and grants that is standard for performance appraisal today.

ACADEMIC SELF-REGULATION AND VOCATIONAL AUTONOMY ARE REPLACED WITH CONTROL BY AUDIT AND MANAGEMENT.

This strategy would have been risk-free. If the investigation had cleared Macchiarini's name, everyone would have benefited — the KI, the Karolinska University Hospital, Hamsten, Macchiarini, Swedish advanced clinical treatment, the international community of regenerative medicine and more. Should it have proved fraud by Macchiarini, and had the rector then taken swift action to terminate his contract, everyone, apart from the one who committed the fraud, would likewise have benefited.

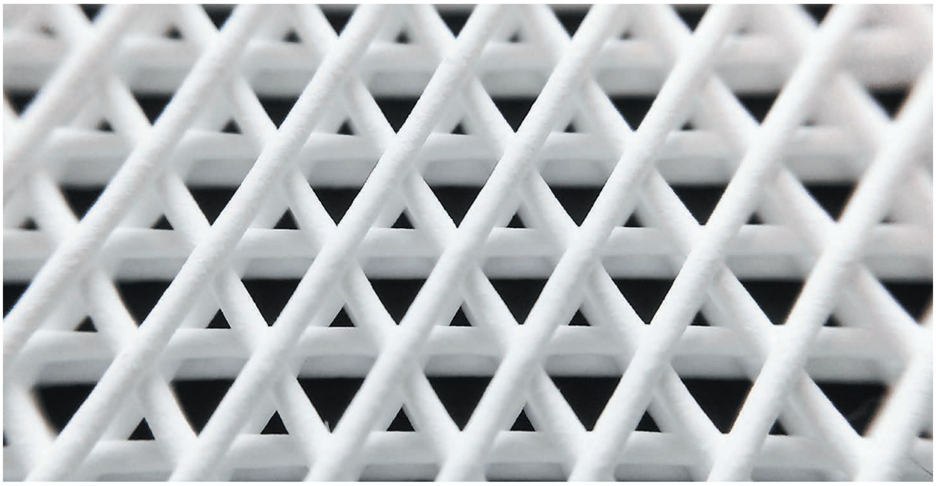
Proper regard to peer review might well have prevented Macchiarini's rise to a prestigious position at the Karolinska. It would, sadly, perhaps not have prevented the deaths of his patients, but it would have avoided the exposure of important institutions such as the KI, the Karolinska University Hospital, and even the Nobel prize, to the crisis of confidence that they are currently experiencing.

As science celebrates its achievements this week, it should remember and cherish the system that produced them. The Karolinska scandal puts the spotlight on the adverse consequences of academic capitalism, which has robbed that system of important safety nets. A return to proper practice is needed to avoid the reputations of other important institutions suffering in the same way in the future. ■

Olof Hallonsten *is a sociologist of science at Lund University, Sweden.*
e-mail: olof.hallonsten@fek.lu.se

RESEARCH HIGHLIGHTS

Selections from the scientific literature



ADAM E. JAKUS

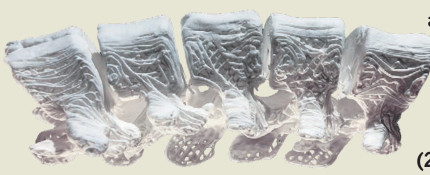
BIOMATERIALS

'Bones' made with 3D printer

Synthetic bones promote natural bone regeneration after being implanted into animals.

Ramille Shah at Northwestern University in Evanston, Illinois, and her colleagues used a 3D printer to generate 'hyperelastic bone'. The main component of the material was hydroxyapatite — a calcium mineral similar to one found in bone — which was mixed with one of two polymers used in medicine and tissue engineering. Grafts built with the material (main picture) and implanted into mice, rats and one macaque became integrated into tissue and stimulated bone growth without adverse effects. Moreover, a 3D-printed 'bone' shaped like a section of human femur was able to withstand loads similar to those experienced naturally.

The material can be rapidly printed into a variety of shapes (human spinal section, inset) and is easy to use in surgery, the authors say.



ADAM E. JAKUS

Sci. Transl. Med. 8, **358ra127 (2016)**

ASTRONOMY

Magnetism drives star birth

Magnetic fields regulate how stars are born from massive clouds of interstellar gas.

A team led by Francesco Fontani at the Arcetri Astrophysical Observatory in Florence, Italy, used high-resolution data from the Atacama Large Millimeter/submillimeter Array telescope in northern Chile to create detailed maps of a particular gas cloud. They found that the gas collapsed under the force of gravity and fragmented, forming a string of clumps that aligned themselves with the magnetic field. The clumps will eventually form the cores of future stars.

The study's findings confirm theoretical predictions that magnetic fields play a major part in where proto-stars form.

Astron. Astrophys. 593, **L14 (2016)**

NEUROSCIENCE

Hunger overrides other motivations

Hungry mice will seek out food in fearful situations that they would normally avoid, and researchers have pinpointed the neurons in the brain that seem to control this behaviour.

Michael Krashes at the US National Institutes of Health in Bethesda, Maryland, and his colleagues stimulated appetite-regulating neurons in the hypothalami of mice that had recently been fed, and observed their behaviour in various settings. They found that the animals were more willing than non-stimulated ones to enter open, unprotected spaces or areas infused with fox odour in order to obtain food. Hungry or brain-stimulated males also opted to pursue food rather than spend time with a female mouse.

Future studies could reveal how these neurons suppress competing drives such as fear and sociality.

Neuron **http://doi.org/brbf (2016)**

MATERIALS

Graphene oxide is stiff yet bendy

An oxidized form of graphene — single-atom-thick layers of carbon — is extremely flexible, despite also being very resistant to stretching.

Cécile Zakri at the University of Bordeaux in France and her colleagues measured how much layers of graphene oxide resist bending by using X-rays to study how easily natural ripples in the sheet can be flattened. They found that graphene oxide is about 100 times easier to bend than graphene, even though both materials have a resistance to stretching along the plane of the sheet that is comparable to that of steel.

Graphene oxide's unique combination of stiffness and superflexibility makes it a suitable material for applications such as flexible but strong electronics, say the authors.

Proc. Natl Acad. Sci. USA **http://doi.org/bq7k (2016)**

CANCER IMMUNOTHERAPY

Dual action of targeted T cells

Immune cells engineered to attack tumours can also be used to deliver cancer-fighting proteins.

T cells that have been engineered to recognize

tumours have shown promise as treatments for certain blood cancers. Hans-Guido Wendel at the Memorial Sloan-Kettering Cancer Center in New York, Karin Tarte of the French National Institute of Health and Medical Research in Rennes and their colleagues took that engineering a step further. The team found that loss of *HVEM* — a gene that is often mutated in some types of lymphoma — fosters lymphoma development in mice.

Injecting a key domain of the normal HVEM protein directly into mouse lymphoma tumours blocked their growth. The authors then engineered T cells to produce the protein and deliver it to the cancer cells, treatment that prevented tumour growth in a mouse model of lymphoma.

Cell **http://doi.org/brbd (2016)**

ASTRONOMY

How black hole obscures itself

A supermassive black hole at the core of a distant galaxy is hiding in a cloak of its own making.

Supermassive black holes are shrouded by doughnut-shaped rings of gas and dust, but scientists are not sure where these come from. A team led by Jack Gallimore of Bucknell University in Lewisburg, Pennsylvania, used the Atacama Large Millimeter/submillimeter Array in Chile to observe galaxy NGC 1068, 14.4 million parsecs (47 million light years) away. They saw hot, ionized clouds of carbon monoxide gas flying away from the galaxy's black hole in opposite directions.

This suggests that the gas originates from the disk of material swirling around the black hole and is flung off by its spinning magnetic field. The findings could alter theories of how black holes interact with their host galaxies.

Astrophys. J. **829, L7 (2016)**

X. HU ET AL./NATURE COMMUN.

CARDIOVASCULAR BIOLOGY

'Good' fat may cut heart disease

Too much dietary fat is associated with heart disease, but one type of fat could help to combat atherosclerosis.

Ebru Erbay at Bilkent University in Ankara and her colleagues studied a mouse model of atherosclerosis, in which the animals develop fatty plaques in their arteries. The scientists found that mice fed a fatty acid called palmitoleate had smaller plaques than those that did not consume it. The fat seems to reduce the number of inflammatory immune cells called macrophages in the plaques. Palmitoleate also blocks a type of inflammation that is triggered by saturated fat in both mouse and human macrophages.

The effects of palmitoleate supplementation should be tested in humans as a possible preventive measure for heart disease, the authors suggest.

Science Transl. Med. **8, 358ra126 (2016)**

MICROBIOLOGY

Toad probiotic fights fungus

Treatment with a skin microbe protects captive toads against a lethal fungal infection.

Valerie McKenzie at the University of Colorado Boulder and her colleagues compared the skin microbiomes of endangered boreal toads (*Anaxyrus boreas*; **pictured**) reared in captivity with hose of wild ones. They found that the diversity of bacterial strains that inhibit the fungal pathogen Bd (*Batrachochytrium dendrobatidis*) was greatly reduced on the captive toads, and that these animals lost the protective microbes over time. As a result, after nearly eight months in captivity, all toads exposed to the fungus died. In a second experiment, inoculating exposed amphibians with a Bd-inhibiting microbe increased the animals' survival by 40%.

Long-term captivity reduces toads' exposure to beneficial environmental microbes that protect them against Bd and other pathogens, the authors say.

Proc. R. Soc. B **283, 20161553 (2016)**

JOEL SARTORE/NATL GEOGRAPHIC/GETTY

MATERIALS

Shape-shifting gel blooms

A gel has been programmed to change shape on its own, without any external triggers.

Most shape-shifting materials require a shift in conditions — for example, temperature or humidity — to flip between two forms. But Andrey Dobrynin at the University of Akron in Ohio, Sergei Sheiko at the University of North Carolina at Chapel Hill and their team created a polymer hydrogel with two types of crosslink: permanent covalent bonds that allow the material to recover its initial shape after deformation, and hydrogen bonds that temporarily hold it in a different configuration. By varying factors including the speed at which the temporary deformation occurs and the length of time it is held in place, the researchers could control how rapidly the material regained its shape, without the need for a trigger. Using this approach, the team created an artificial flower with individually programmed petals that unfolded in sequence (**pictured**).

Such a material could have applications in devices such as medical implants, the authors say.

Nature Commun. 7, **12919 (2016)**

CONSERVATION

Restored forests ignore history

Forests in central Europe were once dominated by conifers, not the broadleaf trees that restoration efforts have focused on growing.

Péter Szabó at the Institute of Botany of the Czech Academy of Sciences in Brno and his colleagues examined fossil pollen from six sites in the central highland region of the Czech Republic, as well as data from a taxonomic survey conducted between 1787 and 1789. They conclude that spruce had been the dominant forest tree since 7,000 BC. This is at odds with the current restoration practice of growing beech and other broadleaf trees, which have long been assumed to be the native trees of the region.

Historical data should be taken into account when restoring forests, the authors suggest.

Conserv. Biol. **http://doi.org/bq7c (2016)**

➲ NATURE.COM
For the latest research published by *Nature* visit:
www.nature.com/latestresearch

SEVEN DAYS *The news in brief*

SPACE

Rosetta rests

The European Space Agency's Rosetta spacecraft successfully crash-landed on the comet 67P/Churyumov–Gerasimenko on 30 September, in a daring finale to its 12-year mission. The craft sent back a continuous stream of data as it descended 19 kilometres to the comet's surface. The move was designed to get scientists the closest possible images and measurements of dust, gas and plasma from a comet. See page 13 for more.

UN space ambition

The United Nations will launch its first space mission in 2021, aiming to give developing nations an opportunity to conduct space research. The UN Office for Outer Space Affairs (UNOOSA) announced on 27 September at an aeronautics congress in Mexico that it will put a payload on the Dream Chaser spacecraft being developed by the Sierra Nevada Corporation in Sparks, Nevada. UNOOSA said that it will soon start soliciting proposals for payloads to be launched into low-Earth orbit. It aims to select a mission by early 2018.

SASHA MORDOVETS/GETTY

Russia suspends plutonium deal with US

On 3 October, Russian President Vladimir Putin suspended an agreement with the United States that requires each country to dispose of 34 tonnes of weapons-grade plutonium, citing "unfriendly" US actions. Under the 2000 deal, which was reaffirmed in 2010, both countries committed to blending the plutonium into mixed-oxide (MOX) fuel for use in nuclear power plants. Delays and cost overruns at a MOX fuel-fabrication facility at the Savannah River Site in South Carolina, however, prompted the US Department of Energy (DOE) to abandon the idea. Instead, the DOE is proposing to dilute and dispose of the plutonium directly. But Russia had opposed that option, claiming that the plutonium could eventually be recovered.

FACILITIES

Laser launch

The world's most powerful X-ray free-electron laser (XFEL), in Hamburg, Germany, officially launched on 6 October. The €1.2-billion (US$1.3-billion) European XFEL, funded by 11 countries, is entering its test phase. When fully operational, it will accelerate bunches of free electrons to near the speed of light, generating X-ray radiation at 27,000 pulses per second. Scientists will use the radiation to study complex molecules and chemical reactions in unprecedented detail. The facility's 1.7-kilometre superconducting linear accelerator was installed in an underground tunnel last month. If tests go to plan, researchers will be able to apply for instrument time starting next year.

POLICY

Climate deal sealed

The European Union's parliament voted to ratify the 2015 Paris climate deal on 4 October, securing enough backing for the agreement to enter into force. The accord needed the support of 55 nations covering 55% of global greenhouse-gas emissions to do so. The European Union accounts for 12% of global emissions. India (responsible for 4% of emissions) ratified the deal on 2 October. Signed last December in Paris by nearly 200 nations, the accord commits countries to keeping global warming to "well below" 2 °C.

French budget

With one eye on next year's elections, the French government has proposed a generous boost for its Ministry of Higher Education and Research in the draft budget for 2017, released on 28 September. The ministry would get a 3.7% spending hike, bringing its total budget to €23.85 billion (US$27 billion) and its research pot to €7.9 billion. It is the largest increase for 15 years, but some fear that already-promised salary raises for civil servants — including many researchers and university teaching staff — could swallow up much of the increased budget.

Wildlife protection

In a significant step, delegates at the meeting of the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES) in Johannesburg, South Africa, agreed on a motion calling for the closure of all domestic ivory markets. Japan, however, has said the non-binding motion won't apply there. But the congress rejected proposals to give African elephants the highest level of protection available. Other actions at the 12-day meeting, which closed on 5 October, included banning

all trade of pangolins, which are used in Chinese medicine and are some of the world's most trafficked mammals; and boosting protection for thresher sharks, known for their long, whip-like tails.

EVENTS

Artificial pancreas

US regulators have approved the first 'artificial pancreas' — a device that automatically adjusts insulin levels on the basis of blood-sugar levels. The US Food and Drug Administration approved the device, which is made by Medtronic of Dublin, on 28 September to treat type 1 diabetes. The artificial pancreas measures blood sugar every five minutes and relies on an insulin pump to adjust insulin levels accordingly.

'Three-parent' baby

A potential world first in fertility therapy — a baby boy conceived using a controversial mitochondrial-replacement technique that mixes DNA from three people — was reported by *New Scientist* on 27 September. The method, called spindle nuclear transfer, moves the nucleus of an egg cell from a mother with faulty mitochondria to the nucleus-free egg of a healthy donor; this is then fertilized with the father's sperm. The

COURTESY OF NEW HOPE FERTILITY CLINIC

procedure was carried out in Mexico by a team from a US clinic, on behalf of a Jordanian couple. The mother of the baby carries a neurological disease called Leigh's syndrome. But with only sparse information available, the claim has not been verified, and some researchers have questioned the ethics of the procedure. The team, led by John Zhang (**pictured**, with baby), is scheduled to present details on 19 October. The boy was born in April. See go.nature.com/2dphaud for more.

Arctic science

Nations have made a joint pledge to improve collaboration on Arctic research. Science ministers and advisers from more than 20 nations and the European Union, plus representatives from indigenous groups, met at the White House on 28 September for the first Arctic-science ministerial meeting to discuss the rapidly changing polar environment. In a joint statement, the ministers announced several projects, including a five-year drive to create an Arctic observation system, led by Norway; an EU-led project on the Arctic's impacts on Northern Hemisphere weather; and a US-led research network that will harness the power of citizen scientists.

AI super-league

Tech giants Google, Facebook, Amazon, IBM and Microsoft will join forces to create an artificial-intelligence (AI) consortium to promote public understanding of the field. The Partnership on Artificial Intelligence to Benefit People and Society, announced on 28 September, will recommend best practices, consult with academics on how AI might affect society, and propose standards for future AI researchers. But two big names are so far conspicuously absent from the group: Apple and Elon Musk's research-focused company OpenAI.

Entangled whales

Two North Atlantic right whales (*Eubalaena glacialis*) were found dead and a third became entangled in fishing gear off the coasts of Maine and Massachusetts between 22 and 24 September. The species, which is endangered, has a population of about 500 in the region. Officials with the US National Oceanic and Atmospheric Administration removed buoys and more than 60 metres of rope from the entangled whale, an 8-year-old female, before she became uncooperative. A necropsy of one of the dead whales revealed that it had died of stress after being entangled in fishing gear.

eLife to charge

The open-access journal *eLife* announced on 29 September that it is dropping one of its most distinctive features: free publishing. From 2017, it will charge a fee of US$2,500 for each accepted paper. The journal, which launched in 2012, has until now had its expenses covered by three of the world's largest private research funders. But it needs another revenue stream to support its business as the number of papers that it receives increases, says its director. The fee is in the range charged by other open-access journals. See go.nature.com/2dwl1hy for more.

AWARDS

Nobel prizes

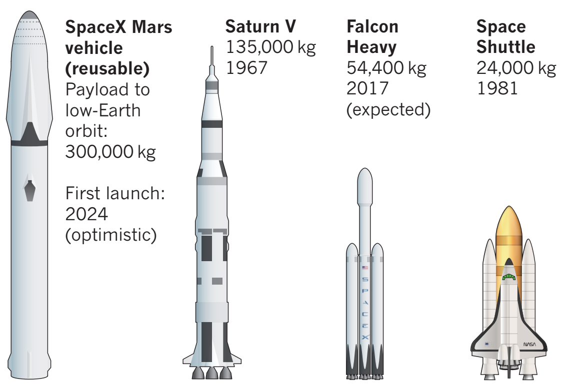
Molecular biologist Yoshinori Ohsumi won the 2016 Nobel Prize in Physiology or Medicine for his work in the field of autophagy: the processes by which the cell digests and recycles its own components. The physics prize was awarded to David Thouless, Duncan Haldane and Michael Kosterlitz for discoveries of exotic behaviour in matter, and for using the mathematics of topology to explain the phenomena. A member of the Nobel physics committee used a bagel and pretzel to aid his explanation of the work (see page 18). *Nature* went to press before the chemistry prize was awarded, but full details are available at go.nature.com/2dnp5bb.

TREND WATCH

SpaceX head Elon Musk has unveiled a plan to colonize Mars. In his yet-to-be-built Interplanetary Transport System, a spaceship designed to carry at least 100 people would be mounted on the most powerful rocket ever built. Both elements are intended to be reusable. After launch, the rocket booster separates in orbit and lands back on Earth. The spaceship, parked in orbit, waits for the booster to return and refuel it with methane and oxygen. Once fully fuelled, the spaceship heads to Mars.

YOU'RE GOING TO NEED A BIGGER ROCKET

At 122 metres, SpaceX's Mars vehicle would be the biggest space-flight system ever built. It is designed to lift into low-Earth orbit more than twice what was possible with NASA's Saturn V Moon rocket.



SOURCE: SPACEX/NATURE

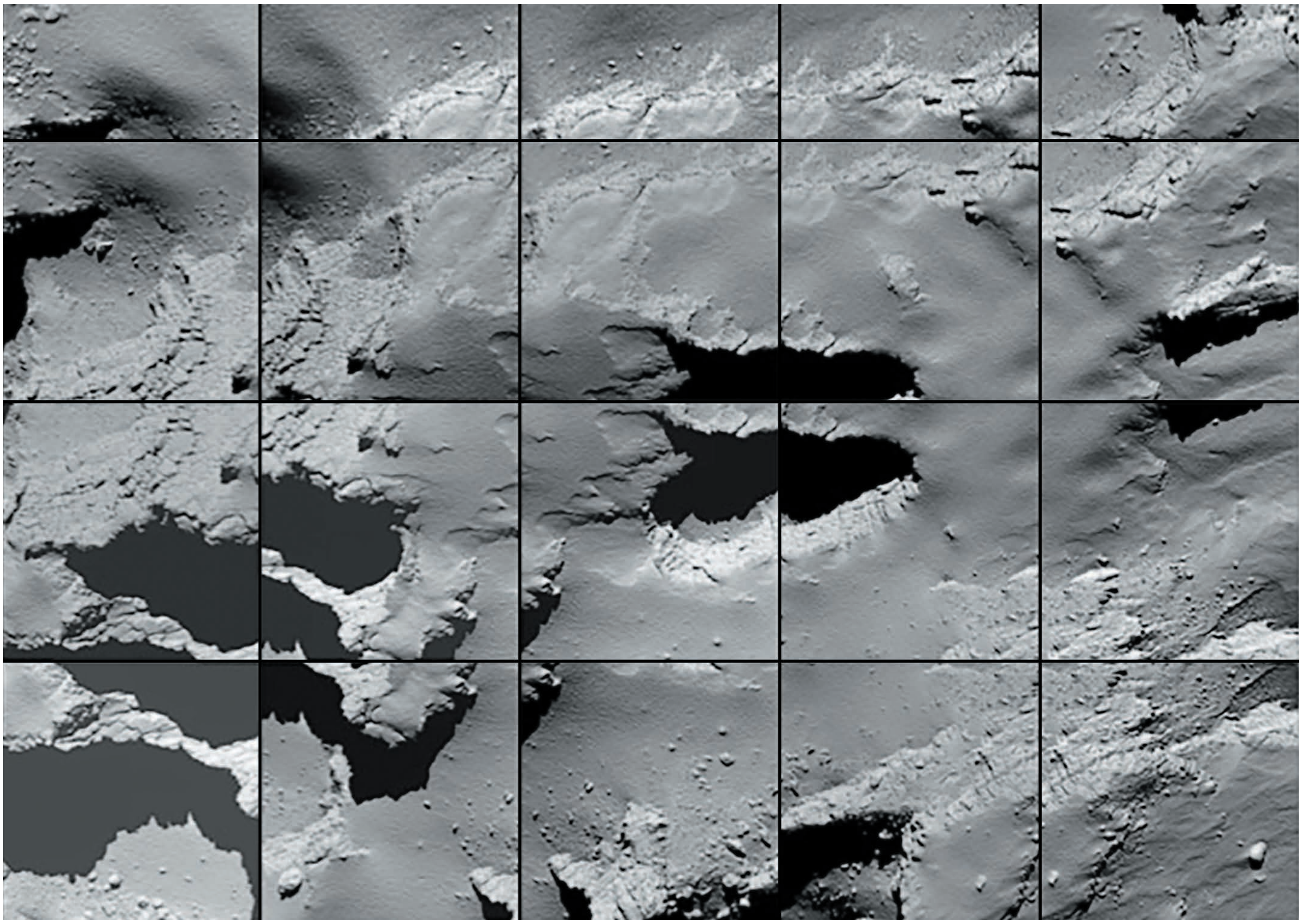
NEWS IN FOCUS

NUCLEAR PHYSICS Two of three US fusion reactors go out of action **p.14**

CONFLICT Syrian seed bank duplicated in Lebanon and Morocco **p.16**

NOBEL PRIZE Studies of cellular recycling and exotic matter snag awards **p.18**

GEOGRAPHY Matthew Hansen maps global forest change from space **p.24**



ESA/ROSETTA/MPS FOR OSIRIS TEAM MPS/UPD/LAM/IAA/SSO/INTA/UPM/DASP/IDA

A sequence of images captured by Rosetta during its descent to the surface of comet 67P.

SPACE SCIENCE

Rosetta crashes into comet

Craft sends back wealth of images in 19-kilometre descent.

BY ELIZABETH GIBNEY

The European Space Agency's comet-orbiting Rosetta spacecraft was successful to the last. It crash-landed on the comet 67P/Churyumov–Gerasimenko within one minute of its scheduled impact time, confirmed at 11:19 UTC on 30 September, ending its 12-year, €1.3-billion (US$1.45-billion) mission with a bump — and a final tranche of data.

The orbiter's crash site has been named Sais, after the site in Egypt where the mission's namesake, the Rosetta stone, was originally displayed. "We can finally say Rosetta has come home to Sais," said mission manager Patrick Martin, speaking from the control room at the European Space Operations Centre (ESOC) in Darmstadt, Germany. "Farewell Rosetta, you've done the job. That was space science at its best."

Flight engineers at ESOC watched quietly for Rosetta's communications signal to flatline — the sign that the craft had landed. At the crucial moment, onlookers looked stunned before breaking into applause to celebrate the culmination of the mission. ▶

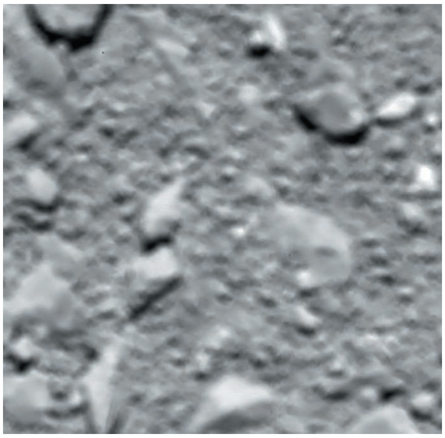
The daring finale was designed to get scientists the closest possible images and measurements of dust, gas and plasma from a comet. Rosetta sent back a continuous stream of data as it drifted down at a sedate walking pace from a height of 19 kilometres onto comet 67P's surface; ESA broadcast the images throughout the descent.

Holger Sierks, principal investigator for Rosetta's OSIRIS instrument (Optical, Spectroscopic, and Infrared Remote Imaging System), showed off the final pictures. A gravel field strewn with pebbles and boulder-like shapes is visible in the crude, unprocessed images. "This will keep us busy," he said.

The craft sent its closest shot just 10 seconds before impact, around 20 metres away from the comet. "That image was extraordinary," says Stephen Lowry, a cometary scientist at the University of Kent in Canterbury, UK, and a member of the OSIRIS camera team.

Rosetta's ultraviolet spectrograph, which studies the characteristic fingerprints in reflected light that reveal the comet's make-up, gathered its last data just minutes before the crash. Alan Stern, a planetary scientist at the Southwest Research Institute in Boulder, Colorado, and principal investigator of the NASA instrument, called the data's 3-metre resolution "unprecedented for ultraviolet studies of comets".

In the coming days, ESOC will use housekeeping data to reconstruct Rosetta's last journey. Estimates suggest that the landing was as close as 40 metres to the target site, with instruments sending back data well within a minute of the crash, says Martin. "The plan worked well until the end, really flawlessly," he says. Most of Rosetta's operations and science staff will now move on to other projects, but Martin will remain on the mission for three years, largely to archive data.



Rosetta's last image of comet 67P/Churyumov–Gerasimenko, taken from about 20 metres up.

ESA/ROSETTA/MPS FOR OSIRIS TEAM MPS/UPD/LAM/IAA/SSO/INTA/UPM/DASP/IDA

So far, scientists have analysed only around 5% of the data that Rosetta has gathered since it began orbiting 67P two years ago, said André Bieler, a planetary scientist at the University of Berne and a member of Rosetta's ROSINA (Rosetta Orbiter Spectrometer for Ion and Neutral Analysis) team, at a meeting at ESOC on the eve of the crash. "We have collected data we haven't had time to look at, but they're there, and they're ready to be assembled," he said.

Rosetta has already made striking findings, including the discovery of water from comet 67P with a different isotopic composition to that on Earth, as well as the presence of molecular oxygen and nitrogen, which points to the comet being as old as the Solar System itself. Scientists also determined how 67P got its strange rubber-duck shape, deducing that the head and body were formed separately.

But many questions remain. A big challenge will be to work out how the pebbles visible in Rosetta's final shots were created, says Lowry. They could have been shaped by dust, which is tossed into the air by sublimating ice and then falls back to the surface. Another tantalizing possibility is that the pebbles are the building blocks from which 67P was originally built. If so, they might be able to tell scientists about the origins of the Solar System.

The Rosetta mission was the first to orbit (rather than just visit) a comet; the first to land a probe on a comet; and the first to conclude with a controlled comet crash-landing. (In 2001, NASA's NEAR Shoemaker mission also crash-landed — but that was on an asteroid, a body that is much larger and nearer to Earth than 67P.) "Rosetta has entered the history books once again," said Johann-Dietrich Wörner, ESA's director-general.

The ability to observe a cometary body changing over time, and from such close quarters, is likely to mean "a true revolution" in cometary science, says Geraint Jones, a planetary scientist at University College London. "It's just a wealth of data. The level of detail is incredible," he says. ■

NUCLEAR PHYSICS

US left with just one working fusion reactor — for now

Design flaw may have doomed machine at Princeton Plasma Physics Lab.

BY JEFF TOLLEFSON

A tough year just got tougher for US fusion researchers. The country's flagship experimental fusion reactor has broken down, less than a year after completing a 4-year, US$94-million upgrade. Now officials at the Princeton Plasma Physics Laboratory (PPPL) in New Jersey are investigating whether problems encountered during fabrication of a key component caused the reactor to fail.

Lab officials say that the machine could be offline for up to a year. Making matters worse, one of the other two fusion reactors funded by the US Department of Energy (DOE) was scheduled to shut down on 30 September. That leaves US scientists with just one major facility to conduct fusion experiments, at the defence contractor General Atomics in San Diego, California.

"It's definitely a challenge for everybody," says Earl Marmar, who oversees the Alcator C-Mod reactor at the Massachusetts Institute of Technology in Cambridge that is shutting down after more than two decades. "We won't be completely without access to experimental facilities, but it's definitely not as good as it could have been for the coming year."

The upgraded Princeton reactor, called the National Spherical Torus Experiment Upgrade (NSTX-U), is twice as powerful as its predecessor. Like other 'tokamak' reactors, including the international ITER project under construction in France, the spherical machine uses magnetic fields to confine a hydrogen plasma. That plasma is then heated until the atoms fuse and release energy. In theory, fusion could power the world indefinitely — and cleanly.

The Princeton machine's breakdown came to light on 27 September, after PPPL director Stewart Prager resigned. Laboratory officials

say that the upgraded reactor started operating at low power in December 2015 and produced 10 weeks of high-quality data. Scientists shut it down in July after discovering that one of the coils that creates the electromagnetic trap was malfunctioning.

Prager says he was thinking about stepping down as director before the reactor coil broke. He elected to depart now, after eight years, so that new leadership can carry the investigation forward and repair the machine. "It's sort of a normal passing of the baton," he says.

PPPL officials initially declined to speculate about the cause of the coil malfunction, saying that an investigation is under way. But the lab later confirmed to *Nature* that questions about the strength of the copper in the faulty coil arose, and were investigated, when the part was being fabricated.

That fact that these concerns arose during the tokamak upgrade suggests that a more careful analysis could have prevented the reactor failure, says Stephen Dean, president of Fusion Power Associates, an advocacy group in Gaithersburg, Maryland. "Mistakes like this do sometimes get made, but with all of the experience the fusion programme has, it should not have happened this way."

PPPL

The experimental fusion reactor at the Princeton Plasma Physics Laboratory is shaped like a cored apple.

HUNTING FOR CLUES

NSTX-U programme director Jonathan Menard says that the finished coil met the laboratory's specifications. He adds that it is not clear whether the part's design or the manufacturing process caused problems. Another coil in the reactor, of a similar design and fabricated from the same grade of copper, has functioned well. The laboratory is planning to replace it nonetheless.

A former researcher at the Princeton laboratory, who declined to be named because he is not authorized to speak about the issue, says that the copper in the faulty coil might have been stronger than it needed to be. That would have made it harder to bend the metal into the desired shape. Even tiny faults in fabrication can cause problems when energy is coursing through the reactor, heating up the coils.

Menard says that after the coil malfunctioned, X-ray analyses found structural anomalies that may have resulted from internal melting when the reactor was operating. PPPL scientists plan to cut the coil open for further investigations. "We are going to have to wait for those results to make a more definitive statement," he says.

Officials aren't sure how much it will cost to repair the reactor, but say that it could take up to a year to bring it back online. Because the fusion reactor was already scheduled to halt operations in late 2016 for six months of maintenance, the net loss of research time may wind up being about six months.

"It's not a good situation for our scientists to only have one machine running."

The breakdown's impacts could extend well beyond the Princeton lab. Marmar had planned to shift people to the Princeton facility once MIT's Alcator reactor shut down. Now, MIT researchers will help Princeton to restart its reactor — and try to conduct their previously planned research by collaborating with teams at General Atomics' reactor and facilities in other countries.

The DOE decided several years ago to close the MIT reactor, but to maintain facilities in Princeton and San Diego. The US Congress reversed that decision once, in 2014, but the US government's 2016 budget assumes that the MIT reactor will shut down.

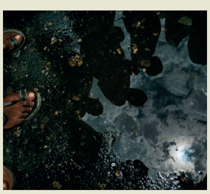
The DOE says that the US fusion-research programme remains on a solid footing, with extensive international partnerships, and will be back at full strength once the Princeton machine returns to service. Others are concerned about how researchers will cope with only one major US reactor in operation.

Dean thinks that the agency ought to keep Alcator C-Mod running for another year, until the Princeton reactor is fixed. "It's not a good situation for our scientists to only have one machine running," he says.

Marmar is ready to restart the MIT reactor if the DOE changes its mind. "The C-Mod facility is planned to be put into a safe shutdown state," he says, "but if desired, could be brought back into service on short notice to support the US and international fusion community." ■

MORE ONLINE

IMAGES OF THE MONTH



Drone lights, cameras and sucking up to a whale go.nature.com/2desclw

DANIEL HAYDUK/AFP/GETTY

MORE NEWS

- 'Three-parent baby' claim raises hopes — and ethical concerns go.nature.com/2dphaud
- Biologists plan scoring system for antibodies go.nature.com/2cp41qg
- Women postdocs less likely to get a glowing reference go.nature.com/2dt0yey

NATURE PODCAST

A limit to lifespan, AI's black-box problem, and ageing stem cells nature.com/nature/podcast

HUSSEIN MALLA/AP

Lebanon's Bekaa valley offers a wealth of ecosystems — and now hosts a growing ICARDA seed bank.

PLANT GENETICS

Syrian seeds get new home

Ancient plant genes will be accessible to scientists again.

BY SHAONI BHATTACHARYA

A major seed bank in Aleppo, Syria, holds genes that might help researchers breed crops to survive climate change. But the conflict tearing the country apart has rendered the bank largely inaccessible for the past four years. Now an effort to duplicate its seed collection at more-accessible locations is ramping up.

On 29 September, the International Center for Agricultural Research in the Dry Areas (ICARDA), which runs the bank in Aleppo, officially launched a sister bank in Terbol, Lebanon, which now hosts 30,000 duplicates. Together with a new bank in Rabat, Morocco, it will make thousands of seeds available to researchers.

"The situation in Syria did not allow us to continue our core activities," says Ahmed Amri, head of genetic resources at ICARDA's research station in Rabat. "I'm happy that we [ICARDA] have established ourselves back to normal."

Seed banks function as bank accounts for plant genes. Collectors deposit seeds, which can later be 'withdrawn' to replenish crops lost in conflict or disaster, to breed new traits into crops — such as pest or heat resistance — and to research the evolution of plants over the ages.

ICARDA's collection, previously held entirely at the bank in Aleppo, is especially valuable because it aims to collect seeds from the world's dry regions. That includes the Fertile Crescent, which spans parts of North Africa, the Middle East, the Caucasus and west Asia, and is thought of as the birthplace of modern agriculture. The collection contains many wild relatives of modern crops such as wheat, barley, lentils and grass pea.

The centre provides researchers and breeders with an average of about 20,000 samples each year, says Amri, with most material going to the United States, to institutions in the nation's breadbasket such as Kansas State University and North Dakota State University. Many wild varieties from arid regions have traits that may help crops to meet the challenges posed by climate change, including resistance to drought, heat and pests, and adaptations to salinity.

ICARDA's gene bank harbours wheat seeds that are the product of thousands of years of adaptation and natural selection, says Maricelis Acevedo, associate director for science for the Delivering Genetic Gains in Wheat project at Cornell University in Ithaca, New York. "Only a small amount of wheat genetic diversity has been utilized and explored."

Although most staff left ICARDA's Aleppo site in 2012, the vault there is intact, according to the last inspection three months ago. But seeds can no longer be moved in or out easily.

Almost all of the seeds in ICARDA's bank have previously been duplicated and sent to banks elsewhere, mainly to the super-secure Svalbard Global Seed Vault in Norway — a.k.a. the 'doomsday vault' — which was set up to provide back-up copies of seeds held in banks worldwide. But this trove is not easily available to scientists. By contrast, ICARDA's collection is mainly meant to be 'active': in other words, available to farmers, researchers and breeders.

In 2015, ICARDA made its first withdrawal of seeds from the Svalbard bank and is now using them to build up stocks in Terbol and Rabat. It will return the stocks to Svalbard and withdraw several more batches to reconstruct the entire Aleppo collection.

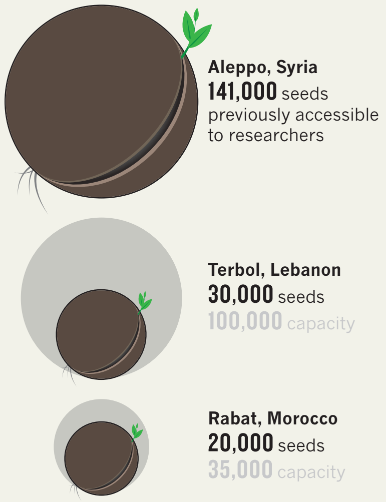
Duplicating the collection in more-accessible gene banks is vital, says Mogens Hovmøller, a plant pathologist at the University of Aarhus in Denmark, who also leads the Global Rust Reference Center. That project was co-founded by ICARDA and is part of an effort to minimize the world's vulnerability to devastating wheat-rust diseases.

The choice of Terbol as a location is a "brilliant move", says Michiel van Slagaren, who worked for ICARDA from 1988 to 1994 and is now at the Kew Royal Botanic Gardens site in Wakehurst, UK. Terbol lies in Lebanon's Bekaa valley, which provides a gradient of conditions from semi-desert to high-rainfall areas

SOURCE: ICARDA

CAUGHT IN CONFLICT

Two seed banks are duplicating a now-inaccessible collection in Aleppo, Syria — but don't have the capacity to host the whole thing.



and so is ideal for testing how seeds grow in different ecosystems, he says.

But the move may also bring risks. The gene bank looks out on the Anti-Lebanon mountain range that forms much of Lebanon's border with Syria and is not far from the conflict. The Bekaa valley also hosts refugees fleeing the civil war.

Van Slageren ponders the potential for the conflict to spill into Lebanon. "You do have to wonder how their minds have been put at ease," he says. He notes that when ICARDA was set up in 1977, its headquarters were in Lebanon, but moved to Syria because of the Lebanese Civil War.

The latest move has also posed staff challenges. Many long-serving members were already close to retirement when ICARDA left Syria, says Amri, and so did not move to Terbol. And funding remains an issue, although ICARDA received significant financial help with the move from various agencies, including the CGIAR Consortium, a global partnership aimed at alleviating poverty and hunger.

The current capacities of the banks in Terbol and Rabat — 100,000 and 35,000, respectively — do not add up to enough to duplicate all 141,000 seeds, representing some 700 species, that Aleppo holds, let alone take on new seeds (see 'Caught in conflict').

Amri is confident. Among other things, previous unrest in Lebanon did not disrupt ICARDA's Terbol station. "It's gone through 20 years of fighting, and we never had any problems," he says. Still, the Moroccan talks wistfully of his years working in Syria. "We enjoyed our lives in Aleppo. It was one of the nicest places to live — wonderful people and a good environment for research at ICARDA." ■

ETHICS

CRISPR concerns

UK bioethics panel eyes the implications of gene editing.

BY HEIDI LEDFORD

From designer babies to engineered mosquitoes, advances in genome-editing technologies such as CRISPR–Cas9 have raised the possibility of tremendous scientific advances — and serious ethical concerns.

In a preliminary 130-page report released on 30 September, the influential London-based Nuffield Council on Bioethics announced that two applications of the technology demand further attention: genome editing in human embryos and in livestock.

It will probably be years before genome editing is used in human reproduction, but it is clear from speaking to scholars and the public that ethical concerns about edited human embryos are at the forefront of many minds, says Karen Yeung, a legal scholar at King's College London and a member of the Nuffield working group. "Human reproductive applications are perhaps the most talked about or controversial area."

The revelation last year that researchers had used CRISPR–Cas9 in human embryos turned a public spotlight on gene editing's potential applications in human reproduction. That study used non-viable embryos for research purposes only (P. Liang *et al. Protein Cell* **6,** 363–372; 2015), but it launched a public debate about whether and how such technologies should be deployed in people.

It also sparked a spate of soul-searching at national academies and agencies around the world. The US National Academies of Sciences, Engineering and Medicine are compiling a report — due in early 2017 — on human applications of genome editing. And an independent group of European ethicists is speaking to the European Commission about forming a steering committee to ensure that CRISPR methods are safe and reliable before being used for medical purposes.

The Nuffield Council also aims to finish its report on ethical questions in human reproduction in early 2017. The working group will focus on the implications of using gene editing to address genetic diseases, says Yeung. Such applications are years away, she says, but are important enough to warrant an early focus. Tinkering with embryos destined to be implanted is against UK law, she notes. If the group finds strong moral arguments in favour of using genome editing to prevent disease, it could take a long time to change that regulation.

The working group would also have to wrestle with drawing the line between ethically acceptable and unacceptable uses, Yeung says.

That discussion is particularly important, says Alta Charo, who studies law and ethics at the University of Wisconsin–Madison. Scientists and ethicists usually focus on serious genetic disorders, but the public conversation often wanders into murkier territory, such as intelligence augmentation. "The lay press tends to do all of these covers about designer babies," she says. "They tend to focus on the things that are the least likely to be genetically determined, but capture our imaginations the most."

VINCENT J. MUSI/NGC

Cattle could be subject to gene editing — one topic being considered by a UK bioethics group.

Use of the technology in livestock comes with issues of its own. These include concerns about animal welfare, and whether and how meat from such animals should be labelled. Labelling is a particularly vexing issue, given that gene-edited animals can be indistinguishable from their natural counterparts with the same mutation.

"Labelling and classification depend on traceability," says John Dupré, a philosopher of science at the University of Exeter, UK, and a member of the Nuffield working group. "Genome editing makes analytical verification of this difficult or impossible."

But some edited livestock — including cattle without horns and pigs that are resistant to disease — are already under development. And the working group felt that there had been comparatively little public discussion of the matter, says Peter Mills, assistant director of the Nuffield Council. "In the livestock, the technology there is pretty much ready to go," he says. "That was something from our point of view that needs to be brought to public attention." ■

PHYSICS

Nobel for 2D exotic matter

Physics award goes to theorists who used topology to explain strange phenomena.

BY ELIZABETH GIBNEY AND DAVIDE CASTELVECCHI

Physics prizewinners Michael Kosterlitz (left), David Thouless (centre) and Duncan Haldane (right).

LEHTIKUVA/RONI REKOMAA/REUTERS; TRINITY HALL, UNIV. CAMBRIDGE; DOMINIC REUTER/REUTERS

David Thouless, Duncan Haldane and Michael Kosterlitz have won the 2016 Nobel Prize in Physics for their theoretical explanations of strange states of matter in 2D materials, known as topological phases.

The British-born trio's work in the 1970s and 1980s laid the foundations for predicting and explaining bizarre behaviours that experimentalists discovered at the surfaces of materials, and inside extremely thin layers. These include superconductivity — the ability to conduct without resistance — and magnetism in very thin materials. At the time, these mathematical theories were quite abstract, said Haldane in an interview with the Nobel Committee just after winning the prize. He said that he was "very surprised and very gratified" to receive the award. But physicists are now exploring similar states of matter for potential use in a new generation of electronics, and in quantum computers.

Thouless and Kosterlitz's breakthroughs began while at the University of Birmingham, UK. The pair demonstrated that, in theory, superconductivity could occur at low temperatures in thin layers of materials, but would disappear at higher temperatures. They also explained the mechanism that would make the effect vanish. Their theory, the Kosterlitz–Thouless (KT) transition, turned out to apply to many different kinds of 2D material.

In 1982, Thouless also explained a phenomenon known as the quantum Hall effect. In this odd effect, when electrons are confined to thin films, chilled to near absolute zero and subjected to a strong magnetic field, they flow in an orderly way with conductivity that increases in steps with an increasing magnetic field. Thouless viewed the problem through the concept of topology, which describes properties that remain unchanged if an object is deformed but not torn. Just as a knot tied in an unbroken circle of string cannot be removed without cutting the string, topological properties tend to be robust. Changes happen only in sudden steps rather than smoothly, and Thouless showed that the quantum Hall effect was just such a topological phenomenon.

Haldane applied the concept of topology to chains of magnetic atoms. These atoms have a quantum property known as spin, and in 1982, he predicted that certain chains of the atoms could show topological properties that result in half spins at either end. Because this quantum property depends on the collective action of the whole chain, rather than on any individual particle, similar phenomena are now being explored as robust ways to encode information in a quantum computer.

"In different ways, they showed how the concept of topology could give rise to new forms of matter that hadn't previously been understood," says Nigel Cooper, a theoretical physicist at the University of Cambridge, UK.

The theorists now all work in the United States: Thouless at the University of Washington, Seattle; Kosterlitz at Brown University in Providence, Rhode Island; and Haldane at Princeton University in New Jersey. Thouless takes half the prize; the other half is split between Kosterlitz and Haldane. ■

NOBEL PRIZES

Medical award for cell recycling

Japanese biologist Yoshinori Ohsumi recognized for work on crucial biological process.

BY RICHARD VAN NOORDEN AND HEIDI LEDFORD

Molecular biologist Yoshinori Ohsumi has won the 2016 Nobel Prize in Physiology or Medicine for his work on autophagy: the processes by which the cell digests and recycles its own components.

The 71-year-old Ohsumi, a professor at the Tokyo Institute of Technology in Yokohama, was recognized for experiments in the 1990s that used baker's yeast (*Saccharomyces cerevisiae*) to identify genes that control how cells destroy their own contents. Similar mechanisms operate in human cells and are sometimes involved in genetic disease.

"He's a very humble yeast geneticist who basically transformed the field," says Sharon Tooze, a cell biologist at the Francis Crick Institute in London.

The term autophagy — from the Greek for 'self-eating' — was coined in 1963 by the Belgian biochemist Christian de Duve, who saw how cells broke down their parts inside a waste-processing sac that he called a lysosome. Biologists now understand that this process is fundamentally important to living cells.

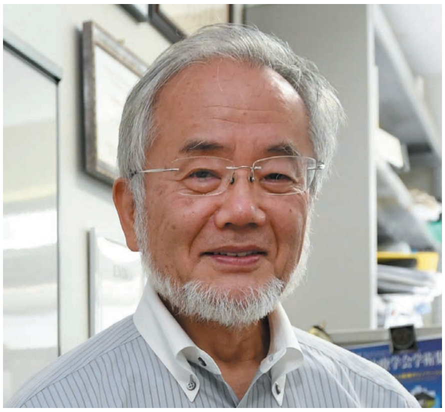
"Without autophagy, our cells won't survive," says Juleen Zierath, a physiologist at the Karolinska Institute in Stockholm who is on the selection committee for the medicine Nobel. When cells are starved, they can consume their own proteins for fuel. The same process can be used to clear out debris such as damaged proteins and organelles, or to ward off invading bacteria and viruses.

SLEEPY BACKWATER

When Ohsumi first started studying autophagy in 1988, "it was kind of a sleepy backwater of a research topic", says biochemist Michael Hall of the University of Basel in Switzerland. "It was basically considered the garbage-disposal system of the cell — just bulk, non-specific degradation of junk."

Ohsumi would go on to develop the first yeast genetics screen to identify genes involved in autophagy. "You can answer the most basic and important questions about the nature of life through yeasts," he said in an interview published on the Tokyo Institute of Technology's website in 2012. But it was a few years before biologists recognized the importance of the process in physiology and disease.

Interest in the field skyrocketed when, in 1999, Beth Levine (now at the University of Texas Southwestern Medical Center in Dallas) and her colleagues reported that a mammalian autophagy gene could suppress tumour growth. The finding launched widespread efforts to learn more about autophagy's role in cancer.



TOKYO INST. TECHNOL./REUTERS

Yoshinori Ohsumi won the 2016 Nobel Prize in Physiology or Medicine.

Disruptions in autophagy have also been linked to Parkinson's disease, type 2 diabetes and other disorders — and research is ongoing to develop drugs that can affect the process.

Researchers' understanding of the complex role of autophagy in cancer has become more detailed: the process seems to inhibit tumours in the early stages of growth, but can also fuel cancer once it has spread, says Hall.

Ohsumi, who will collect 8 million Swedish kronor (US$940,000) for the Nobel prize, won the ¥50-million (US$626,000) Kyoto Prize in basic sciences in 2012 for his autophagy work.

Others have made key contributions to the field, and were considered contenders for a share of a Nobel. Biochemist Michael Thumm of the University Medical Center Göttingen in Germany also discovered autophagy genes, as did cell biologist Daniel Klionsky of the University of Michigan in Ann Arbor.

"If they're going to give it to just one, Ohsumi's the one," says Hall. "But it also would have been good to include other people."

In Japan, the prize had been widely anticipated for the past few years, with journalists showing up regularly to ask Ohsumi for interviews, says Hitoshi Nakatogawa, a biologist at the Tokyo Institute of Technology who has worked with Ohsumi for a decade. When colleagues heard of Ohsumi's win — around two hours before the official announcement — they gathered to celebrate in his lab. "We talked about how great it was that he won it alone," he says.

"Ohsumi never overlooks anything, even in the most banal kind of experiment," Nakatogawa adds. "He doesn't care about whether it will lead to something useful, whether a breakthrough can be expected, whether it will lead to more funding. He just follows his curiosity." ■

Additional reporting by **David Cyranoski**.

THE BLACK BOX

ILLUSTRATION BY SIMON PRADES

Machine learning is becoming ubiquitous in basic research as well as in industry. But for scientists to trust it, they first need to understand what the machines are doing.

BY DAVIDE CASTELVECCHI

Dean Pomerleau can still remember his first tussle with the black-box problem. The year was 1991, and he was making a pioneering attempt to do something that has now become commonplace in autonomous-vehicle research: teach a computer how to drive.

This meant taking the wheel of a specially equipped Humvee military vehicle and guiding it through city streets, says Pomerleau, who was then a robotics graduate student at Carnegie Mellon University in Pittsburgh, Pennsylvania. With him in the Humvee was a computer that he had programmed to peer through a camera, interpret what was happening out on the road and memorize every move that he made in response. Eventually, Pomerleau hoped, the machine would make enough associations to steer on its own.

On each trip, Pomerleau would train the system for a few minutes, then turn it loose to drive itself. Everything seemed to go well — until one day the Humvee approached a bridge and suddenly swerved to one side. He avoided a crash only by quickly grabbing the wheel and retaking control.

Back in the lab, Pomerleau tried to understand where the computer had gone wrong. “Part of my thesis was to open up the black box and figure out what it was thinking,” he explains. But how? He had programmed the computer to act as a ‘neural network’ — a type of artificial intelligence (AI) that is modelled on the brain, and that promised to be better than standard algorithms at dealing with complex real-world situations. Unfortunately, such networks are also as opaque as the brain. Instead of storing what they have learned in a neat block of digital memory, they diffuse the information in a way that is exceedingly difficult to decipher. Only after extensively testing his software’s responses to various visual stimuli did Pomerleau discover the problem: the network had been using grassy roadsides as a guide to the direction of the road, so the appearance of the bridge confused it.

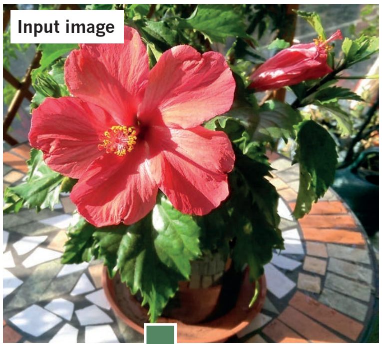
Twenty-five years later, deciphering the black box has become exponentially harder and more urgent. The technology itself has exploded in complexity and application. Pomerleau, who now teaches robotics part-time at Carnegie Mellon, describes his little van-mounted system as “a poor man’s version” of the huge neural networks being implemented on today’s machines. And the technique of deep learning, in which the networks are trained on vast archives of big data, is finding commercial applications that range from self-driving cars to websites that recommend products on the basis of a user’s browsing history.

It promises to become ubiquitous in science, too. Future radio-astronomy observatories will need deep learning to find worthwhile signals in their otherwise unmanageable amounts of data; gravitational-wave detectors will use it to understand and eliminate the tiniest sources of noise; and publishers will use it to scour and tag millions of research papers and books. Eventually, some researchers believe, computers equipped with deep learning may even display imagination and creativity. “You would just throw data at this machine,

DO AIs DREAM OF ELECTRIC SHEEP?

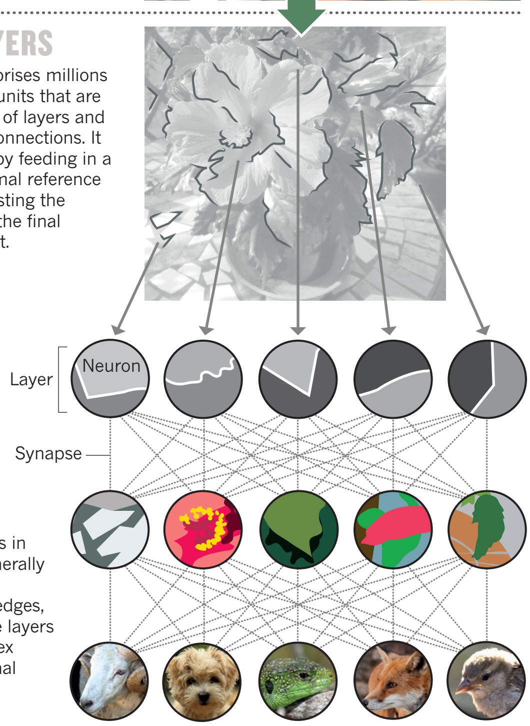
In an effort to understand how artificial neural networks encode information, researchers invented the Deep Dream technique.

Starting with a network (below) that has been trained to recognize shapes such as animal faces, Deep Dream gives it an image of, say, a flower. Then it repeatedly modifies the flower image to maximize the network's animal-face response.

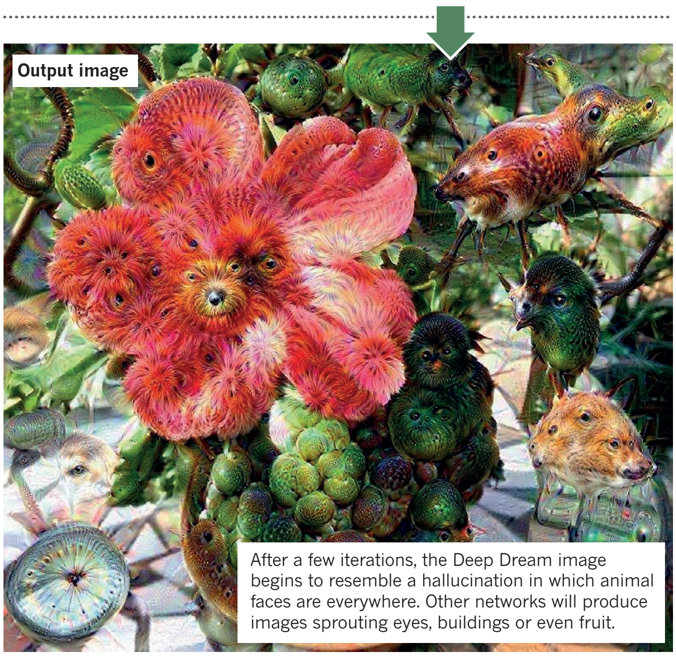


HIDDEN LAYERS

The network comprises millions of computational units that are stacked in dozens of layers and linked by digital connections. It has been trained by feeding in a vast library of animal reference images, then adjusting the connections until the final response is correct.



After training, units in the first layers generally respond to simple features, such as edges, while intermediate layers respond to complex shapes and the final layers respond to complete faces.



After a few iterations, the Deep Dream image begins to resemble a hallucination in which animal faces are everywhere. Other networks will produce images sprouting eyes, buildings or even fruit.

DESIGN: NIK SPENCER/NATURE; PHOTOS: KEITH MCDUFFEE/FLICKR/CC BY; DJHUMSTER/FLICKR/CC BY-SA; BERNARD DUPONT; LINDA STANLEY; PHIL FIDDYMENT/FLICKR/CC BY

▶ and it would come back with the laws of nature," says Jean-Roch Vlimant, a physicist at the California Institute of Technology in Pasadena.

But such advances would make the black-box problem all the more acute. Exactly how is the machine finding those worthwhile signals, for example? And how can anyone be sure that it's right? How far should people be willing to trust deep learning? "I think we are definitely losing ground to these algorithms," says roboticist Hod Lipson at Columbia University in New York City. He compares the situation to meeting an intelligent alien species whose eyes have receptors not just for the primary colours red, green and blue, but also for a fourth colour. It would be very difficult for humans to understand how the alien sees the world, and for the alien to explain it to us, he says. Computers will have similar difficulties explaining things to us, he says. "At some point, it's like explaining Shakespeare to a dog."

Faced with such challenges, AI researchers are responding just as Pomerleau did — by opening up the black box and doing the equivalent of neuroscience to understand the networks inside. Answers are not insight, says Vincenzo Innocente, a physicist at CERN, the European particle-physics laboratory near Geneva, Switzerland who has pioneered the application of AI to the field. "As a scientist," he says, "I am not satisfied with just distinguishing cats from dogs. A scientist wants to be able to say: 'the difference is such and such.'"

GOOD TRIP

The first artificial neural networks were created in the early 1950s, almost as soon as there were computers capable of executing the algorithms. The idea is to simulate small computational units — the 'neurons' — that are arranged in layers connected by a multitude of digital 'synapses' (see 'Do AIs dream of electric sheep?') Each unit in the bottom layer takes in external data, such as pixels in an image, then distributes that information up to some or all of the units in the next layer. Each unit in that second layer then integrates its inputs from the first layer, using a simple mathematical rule, and passes the result further up. Eventually, the top layer yields an answer — by, say, classifying the original picture as a 'cat' or a 'dog'.

The power of such networks stems from their ability to learn. Given a training set of data accompanied by the right answers, they can progressively improve their performance by tweaking the strength of each connection until their top-level outputs are also correct. This process, which simulates how the brain learns by strengthening or weakening synapses, eventually produces a network that can successfully classify new data that were not part of its training set.

That ability to learn was a major attraction for CERN physicists back in the 1990s, when they were among the first to routinely use large-scale neural networks for science: the networks would prove to be an enormous help in reconstructing the trajectories of subatomic shrapnel coming out of particle collisions at CERN's Large Hadron Collider.

But this form of learning is also why information is so diffuse in the network: just as in the brain, memory is encoded in the strength of multiple connections, rather than stored at specific locations, as in a conventional database. "Where is the first digit of your phone number stored in your brain? Probably in a bunch of synapses, probably not too far from the other digits," says Pierre Baldi, a machine-learning researcher at the University of California, Irvine. But there is no well-defined sequence of bits that encodes the number. As a result, says computer scientist Jeff Clune at the University of Wyoming in Laramie, "even though we make these networks, we are no closer to understanding them than we are a human brain".

To scientists who have to deal with big data in their respective disciplines, this makes deep learning a tool to be used with caution. To see why, says Andrea Vedaldi, a computer scientist at the University of Oxford, UK, imagine that in the near future, a deep-learning neural network is trained using old mammograms that have been labelled according to which women went on to develop breast cancer. After this training, says Vedaldi, the tissue of an apparently healthy woman could

already 'look' cancerous to the machine. "The neural network could have implicitly learned to recognize markers — features that we don't know about, but that are predictive of cancer," he says.

But if the machine could not explain how it knew, says Vedaldi, it would present physicians and their patients with serious dilemmas. It's difficult enough for a woman to choose a preventive mastectomy because she has a genetic variant known to substantially up the risk of cancer. But it could be even harder to make that choice without even knowing what the risk factor is — even if the machine making the recommendation happened to be very accurate in its predictions.

"The problem is that the knowledge gets baked into the network, rather than into us," says Michael Tyka, a biophysicist and programmer at Google in Seattle, Washington. "Have we really understood anything? Not really — the network has."

> *"The problem is that the knowledge gets baked into the network, rather than into us."*

Several groups began to look into this black-box problem in 2012. A team led by Geoffrey Hinton, a machine-learning specialist at the University of Toronto in Canada, entered a computer-vision competition and showed for the first time that deep learning's ability to classify photographs from a database of 1.2 million images far surpassed that of any other AI approach[1].

Digging deeper into how this was possible, Vedaldi's group took algorithms that Hinton had developed to improve neural-network training, and essentially ran them in reverse. Rather than teaching a network to give the correct interpretation of an image, the team started with pretrained networks and tried to reconstruct the images that produced them[2]. This helped the researchers to identify how the machine was representing various features — as if they were asking a hypothetical cancer-spotting neural network, 'What part of this mammogram have you decided is a marker of cancer risk?'

Last year, Tyka and fellow Google researchers followed a similar approach to its ultimate conclusion. Their algorithm, which they called Deep Dream, starts from an image — say a flower, or a beach — and modifies it to enhance the response of a particular top-level neuron. If the neuron likes to tag images as birds, for example, the modified picture will start showing birds everywhere. The resulting images evoke LSD trips, with birds emerging from faces, buildings and much more. "I think it's much more like a hallucination" than a dream, says Tyka, who is also an artist. When he and the team saw the potential for others to use the algorithm for creative purposes, they made it available to anyone to download. Within days, Deep Dream was a viral sensation online.

Using techniques that could maximize the response of any neuron, not just the top-level ones, Clune's team discovered in 2014 that the black-box problem might be worse than expected: neural networks are surprisingly easy to fool with images that to people look like random noise, or abstract geometric patterns. For instance, a network might see wiggly lines and classify them as a starfish, or mistake black-and-yellow stripes for a school bus. Moreover, the patterns elicited the same responses in networks that had been trained on different data sets[3].

Researchers have proposed a number of approaches to solving this 'fooling' problem, but so far no general solution has emerged. And that could be dangerous in the real world. An especially frightening scenario, Clune says, is that ill-intentioned hackers could learn to exploit these weaknesses. They could then send a self-driving car veering into a billboard that it thinks is a road, or trick a retina scanner into giving an intruder access to the White House, thinking that the person is Barack Obama. "We have to roll our sleeves up and do hard science, to make machine learning more robust and more intelligent," concludes Clune.

Issues such as these have led some computer scientists to think that deep learning with neural networks should not be the only game in town. Zoubin Ghahramani, a machine-learning researcher at the University of Cambridge, UK, says that if AI is to give answers that humans can easily interpret, "there's a world of problems for which deep learning is just not the answer". One relatively transparent approach with an ability to do science was debuted in 2009 by Lipson and computational biologist Michael Schmidt, then at Cornell University in Ithaca, New York. Their algorithm, called Eureqa, demonstrated that it could rediscover the laws of Newtonian physics simply by watching a relatively simple mechanical object — a system of pendulums — in motion[4].

Starting from a random combination of mathematical building blocks such as +, −, sine and cosine, Eureqa follows a trial-and-error method inspired by Darwinian evolution to modify the terms until it arrives at the formulae that best describe the data. It then proposes experiments to test its models. One of its advantages is simplicity, says Lipson. "A model produced by Eureqa usually has a dozen parameters. A neural network has millions."

ON AUTOPILOT

Last year, Ghahramani published an algorithm that automates the job of a data scientist, from looking at raw data all the way to writing a paper[5]. His software, called Automatic Statistician, spots trends and anomalies in data sets and presents its conclusion, including a detailed explanation of its reasoning. That transparency, Ghahramani says, is "absolutely critical" for applications in science, but it is also important for many commercial applications. For example, he says, in many countries, banks that deny a loan have a legal obligation to say why — something a deep-learning algorithm might not be able to do[5].

Similar concerns apply to a wide range of institutions, points out Ellie Dobson, director of data science at the big-data firm Arundo Analytics in Oslo. If something were to go wrong as a result of setting the UK interest rates, she says, "the Bank of England can't say, 'the black box made me do it'".

Despite these fears, computer scientists contend that efforts at creating transparent AI should be seen as complementary to deep learning, not as a replacement. Some of the transparent techniques may work well on problems that are already described as a set of abstract facts, they say, but are not as good at perception — the process of extracting facts from raw data.

Ultimately, these researchers argue, the complex answers given by machine learning have to be part of science's toolkit because the real world is complex: for phenomena such as the weather or the stock market, a reductionist, synthetic description might not even exist. "There are things we cannot verbalize," says Stéphane Mallat, an applied mathematician at the École Polytechnique in Paris. "When you ask a medical doctor why he diagnosed this or this, he's going to give you some reasons," he says. "But how come it takes 20 years to make a good doctor? Because the information is just not in books."

To Baldi, scientists should embrace deep learning without being "too anal" about the black box. After all, they all carry a black box in their heads. "You use your brain all the time; you trust your brain all the time; and you have no idea how your brain works." ■

Davide Castelvecchi *is a reporter for* Nature *in London.*

1. Krizhevsky, A., Sutskever, I. & Hinton, G. E. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems* 1097–1105 (2012).
2. Mahendran, A. & Vedaldi, A. Preprint at http://arxiv.org/abs/1412.0035 (2014).
3. Nguyen, A., Yosinski, J. & Clune, J. Preprint at https://arxiv.org/abs/1412.1897 (2014).
4. Lipson, H. & Schmidt, M. *Science* **324,** 81–85 (2009).
5. Ghahramani, Z. *Nature* **521,** 452–459 (2015).

The tree counter

Geographer* Matthew Hansen *is creating real-time maps that show where forests are being destroyed. Not everyone believes them.

BY GABRIEL POPKIN

CHRIS MADDALONI/*NATURE*

Ask Matthew Hansen to show off his data and he hunches over his computer like a possessed video gamer. With a few mouse clicks, he flies over the globe and zooms in on a forest in Indonesia. The area is designated as a preserve — supposedly protected from deforestation — but Hansen's data reveal a different reality. Bird's-eye images of the trees taken every eight days flash by on the screen. At first, a few red spots perforate the green canopy around the preserve's edge. Then they spread, like bloodstains. "That's got to be illegal fires," he says. "The forest is getting chewed up."

Hansen is among the world's foremost forest sentries. In 2013, he and his colleagues used satellite data to produce the first global, high-resolution maps of where trees are growing and disappearing[1]. Those images revealed some large-scale patterns for the first time, such as that Indonesia had nearly equalled Brazil as the country with the world's highest rate of tropical deforestation. Since then, his team has refined its methods and can now reveal the loss of trees within days.

Just as important is what Hansen does with the underlying data. Unlike some scientists, he makes them freely available online, giving activists, companies and others the ability to monitor activities such as illegal logging and mining, which have destroyed millions of hectares of forest per year over the past few decades. The data have enabled non-governmental organizations (NGOs) and officials in Peru, Congo and other nations to see deforestation as it happens. And they let countries monitor each other's trees — potentially a crucial step in enforcing the international climate agreement signed in Paris last December.

But some have argued that the maps do not always work as advertised. For instance, they lump together destruction of natural forests and the harvesting of managed ones, which critics say leads to inflated estimates of deforestation. And others question whether satellites can monitor forest loss and growth accurately enough to determine how well countries are complying with their commitments on climate change and deforestation, including the Paris deal.

One thing no one disputes is that Hansen is showing the world how mapping from the sky can have an impact on the ground. "If you want to know what's up, you look at what Matt's doing," says Martin Herold, a remote-sensing expert at Wageningen University in the Netherlands. "Nobody's even close."

WANDERING START

Hansen instantly disarms people with his down-to-earth nature. On an unseasonably warm day earlier this year, he was wearing shorts and a short-sleeved shirt when his assistant reminded him that he was due at a meeting. "I'm not dressed for that at all," he laughed as he set off across the campus of the University of Maryland in College Park. His informality helps when working with both African farmers and Hollywood actors, with whom he mingles as easily as with other scientists and policy wonks. But beneath the casual exterior is an intensity that has made Hansen one of the world's most sought-after experts on forests.

Growing up in Indiana surrounded by farm fields, Hansen did not spend a lot of time among trees. But he was struck by trips to the state's few remaining patches of original hardwood forest, which reminded him of Lothlórien, the sylvan kingdom of the elves in *The Lord of the Rings*. He studied electrical engineering at university and then was accepted into law school, but neither stoked his passion. What did excite him was adventure, and he got plenty of it when he headed to what was then Zaire (now the Democratic Republic of the Congo) to volunteer with the Peace Corps.

But when he returned, he still had no clear career direction. "I came back and I thought, what do I like? I like maps," he says. So he went to the University of North Carolina in Charlotte for master's degrees in geography and civil engineering. He took a job at the University of Maryland in 1994 and has been mapping land-cover change using satellite data ever since, picking up a PhD in 2002.

"HE REALLY HAS AN UNDERSTANDING OF WHAT THIS PLANET'S MADE OF."

Hansen has pursued a single goal: to map global land cover with the highest possible resolution using cheap or free data, to better visualize the human footprint on the planet. He has specialized in writing programs to identify diverse types of vegetation — from boreal conifers to palm plantations — using the handful of light frequencies that satellite sensors collect. "He's an exceptionally good geographer," says long-time colleague Thomas Loveland of the US Geological Survey in Sioux Falls, South Dakota. "He really has an understanding of what this planet's made of."

Hansen and his colleagues also meticulously 'ground-truth' their maps by picking random samples of GPS points and getting to them by any means necessary. "It's his favourite type of vacation, to throw random points on ground and go visit them," says his postdoc Alexandra Tyukavina.

In the mid-1990s, when Hansen was starting, the best information about tree cover came from country-level ground-based assessments, in which crews measured individual trees in representative plots and then extrapolated across large regions. Such measurements were — and still are — used alongside remote-sensing data by the Food and Agriculture Organization of the United Nations (FAO) in its periodic global forest assessments. But many countries lack the resources to conduct regular surveys, and others publish statistics that seem unreliable. So Hansen set his sights on producing what he calls a "globally consistent, locally relevant product" from data available to everyone in the world.

But first he had to wait for technology — sensors in space and computer processing power on the ground — to catch up. The first global land-cover map from the University of Maryland came out[2] in 1994, using data from the Advanced Very High Resolution Radiometer (one of a series of orbiting imagers operated by the US National Oceanic and Atmospheric Administration). It had enormous pixels of one degree latitude by one degree longitude, much too coarse to make out details of forests.

A big step forward came when NASA launched its two Moderate Resolution Imaging Spectroradiometer (MODIS) instruments, which gather data at a resolution of up to 250 metres. In 2008, Hansen and his colleagues produced a map[3] that started to reveal large-scale trends in the tropics, such as that nearly half of widespread humid tropical-forest loss between 2000 and 2005 occurred in Brazil. Around that time, scientists working for both the Brazilian government and local NGOs used MODIS and other data sources to develop their own maps and issue alerts when large clearings appeared. This helped officials to use financial pressure, law enforcement and other means to dramatically reduce deforestation in the Amazon, the world's largest and most carbon-rich tropical-forest region.

That success inspired Hansen. But in many other tropical countries, rising consumer demand for commodities such as cattle, soya beans and palm oil has created powerful incentives to clear tropical forests. And in poorer countries, where heavy tree-felling equipment is rare and clearings tend to be small, MODIS's blocky images have proved less useful. Hansen knew that he needed to make his maps sharp enough to show roads snaking their way into previously untouched forests — an almost universal harbinger of larger clear-cutting. "We had to push the spatial resolution because we're interested in humans," he says.

In fact, the data that he needed already existed. Since 1972, Landsat satellites had been collecting images of Earth's surface, starting at a resolution of 80 by 80 metres per pixel and improving to 30 metres in 1982 — roughly the size of two basketball courts side-by-side. But those images had to be bought individually, at costs from hundreds to thousands of dollars each — much too expensive for a global study.

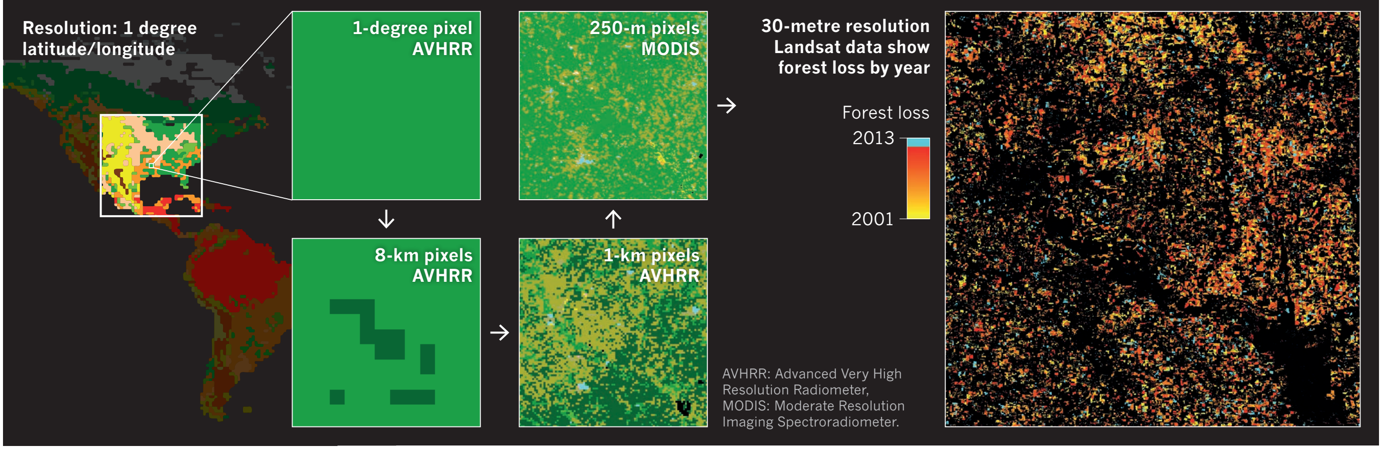
That changed in 2008, when the US government made all Landsat images free, including 3.6 million archived ones. Hansen immediately began making 30-metre-resolution maps showing how tree cover was changing in regions of interest, such as Indonesia and parts of Russia.

But making a global map still required processing power out of reach of any university computer cluster. A solution appeared when Hansen met Google engineer Rebecca Moore at a conference in Brazil. Moore was looking for scientists to try out her Earth Engine, a platform to analyse remote-sensing data using Google's cloud-computing capabilities. Hansen and Moore's teams processed the Landsat archive back to 2000 and translated it into annually updated maps that anybody with a computer and an Internet connection could view. "Matt was the first scientist who really leapt onto the platform with a global-scale analysis," Moore says.

In 2013, Hansen, Moore, Loveland and others published[1] their results in *Science*, showing where trees had appeared or disappeared every year from 2000 to 2012. The maps lit up the research community, which for the first time could see the world's forests shift in one consistent picture (see 'Better eyesight in space'). The fact that Hansen put his raw data on the web for others to scrutinize and use has also drawn admiration.

But it didn't take long for the critics to chime in. Many have objected to

BETTER EYESIGHT IN SPACE | Over the past two decades, Matthew Hansen and his colleagues have used satellite data with successively better resolution to map forests in increasingly fine detail.



SOURCE: M. C. HANSEN

Hansen's use of 'forest', which he defines to include oil-palm plantations and agroforestry, categories not included in FAO data sets. That made his deforestation estimates higher than many previous ones, such as the FAO's. The widespread publicity has further stoked concerns that non-experts are ill-equipped to interpret the data. "I personally think the data set was in some sense oversold," says Herold.

Hansen's visibility added to the scientific scrutiny. On the day that his *Science* paper was published, for example, he was in California showing his maps to actor Harrison Ford in a scene filmed for the 2014 US television series 'Years of Living Dangerously'. Ford later confronted Indonesia's forestry minister with some of the findings.

Other concerns have emerged. Some drier forests, such as those in parts of Africa and South America, have relatively sparse tree cover and might never reach the threshold that Hansen uses to define forest, which is that 30% of a pixel is occupied by vegetation at least 5 metres tall. So when those areas are cleared, the change might not register as forest loss, says Peter Holmgren, director of the Center for International Forestry Research in Bogor, Indonesia. Satellites struggle even more to capture forest gain, he adds, because the signal from growing trees is subtler than that of trees falling. For these and other reasons, he has warned against using Hansen's data to assess progress towards international climate and deforestation commitments, arguing that nations should instead invest in on-the-ground monitoring systems.

Hansen acknowledges that his maps do not supply everything. "You can't fit everybody's needs," he says. But his team is working to add data and make improvements that will show what activities are causing forests to change, and will differentiate plantations from natural forests. "That's what we have to do next, to make it more valuable."

Some of the objections have been more political. Hansen's map was particularly embarrassing for Indonesia because it came out during the 2013 UN climate talks, and revealed that deforestation rates in the country had spiked after a 2011 moratorium on new logging permits was announced. Indonesia's forestry ministry countered that Hansen and his colleagues were including large areas that the government had designated as plantation, unfairly overstating the deforestation.

Hansen's group responded the following year with a more sophisticated analysis[4], which confirmed that, in 2012, more primary tropical forest had fallen in Indonesia than in any other country.

For Hansen, the country's refusal to come clean about its forests is frustrating. But increasing transparency will take time, says Belinda Margono, a scientist with the Indonesian Ministry of Forestry who earned her PhD with Hansen and led the follow-up study by his group. She says that the maps have already helped to set that shift in motion, by promoting a culture of data sharing and openness, and by creating pressure to respond. "Sometimes the government has more courage to release the data after they see what's reported by the global system."

Larger forces are also at work. Nations and corporations are under increasing pressure to show that they are conserving forest to meet commitments under the Paris agreement or in sustainability-certification programmes for products such as palm oil. Since his 2013 paper, Hansen has become a globe-trotting door-to-door salesman of sorts, hawking his maps to forest ministers, corporate accountability officers, NGOs and others who need to keep an eye on forests.

IMMORTALIZED DATA

As almost 200 nations were hammering out the climate deal in Paris last December, Hansen was nearby, receiving a glowing introduction before he spoke at an environmental conference. "Matt and his team ushered in really a new era of measuring deforestation," said Frances Seymour, a forest-policy researcher at the Center for Global Development in Washington DC. "He's now immortalized because everybody talks about the Matt Hansen data on tree-cover change."

Hansen is now working to push his technique even further. Inspired by Brazil's alerts, he has begun processing and displaying data on tree loss as it happens in Peru, Congo, parts of Indonesia and Brazil. In the few months since the alerts went public, Peruvian environmental ministry personnel have used them to expose and shut down an illegal gold-mining operation. The alerts' very existence can have an impact, says remote-sensing scientist Fred Stolle of the World Resources Institute in Washington DC, which is releasing them weekly on its Global Forest Watch online platform. "People know now that they can be seen from space."

Hansen hopes to expand his alerts to the whole tropics by the end of the year, and later to cover the globe. The European Space Agency's Sentinel-2 satellites, which will collect data starting next year with a resolution of up to 10 metres, will enable him to update even more frequently.

Between the travel and the research, Hansen keeps a hectic schedule. But on a rare quiet afternoon, he can explore the world's forests from his desk on the edge of the Maryland campus. As he pans over Peru, a sea of green gives way to a rectangular island of pink that has grown during the past two years. "Someone went out there and clear-cut that," he says.

The view that Hansen has opened up, of trees falling all over the world, does not always reflect the best in people. "It's fucking alarming," he says. "The human footprint is amazing. We are a rapacious species."

But making that view available to everyone, he says, could help to rein our species in. "I hope it will bring some order to the chaos." ■

Gabriel Popkin *is a freelance journalist in Mount Rainier, Maryland.*

1. Hansen, M. C. *et al. Science* **342,** 850–853 (2013).
2. DeFries, R. S. & Townshend, J. R. G. *Int. J. Remote Sens.* **15,** 3567–3586 (1994).
3. Hansen, M. C. *et al. Proc. Natl Acad. Sci. USA* **105,** 9439–9444 (2008).
4. Margono, B. A., Potapov, P. V., Turubanova, S., Stolle, F. & Hansen, M. C. *Nature Clim. Change* **4,** 730–735 (2014).

CHRIS RATCLIFFE/BLOOMBERG/GETTY

Wind turbines near Fjerritslev, Denmark.

Clean up energy innovation

Agree on definitions and baselines to track investments in decarbonizing the world's energy system, urge **Lucien Georgeson**, **Mark Maslin** and **Martyn Poessinouw**.

The Paris climate agreement to keep global average temperature rise below 2 °C requires the world to switch rapidly to low-carbon energy. Global carbon emissions must peak by 2020, fall to zero between 2060 and 2080 and become negative by 2100[1]. The effort and investment needed would be immense, but it could happen: in 1800, the British government spent one-quarter of its per capita expenditure on becoming the world's major naval power[2]; the US Interstate Highway System cost US$560 billion (in 2007 dollars) over 37 years of construction[3].

Clearly, a huge global commitment to clean-energy research and development (R&D) is needed. Two global partnerships were proposed in 2015 to push governments to make the massive investments required: Mission Innovation and the Global Apollo Programme.

Mission Innovation has got countries to pledge to do more R&D on clean energy. But it is not binding and its targets are open to interpretation, being 'bottom up' and voluntary. Global Apollo set narrower 'top down' priorities, but in so doing it has won little national support. Neither covers private spending on R&D, which dwarfs public outlay, is hard to audit and complex to influence.

These initiatives will shape clean-energy research over the next few decades. They need to improve in three respects: their baselines, definitions and private partnerships.

SMOKE AND MIRRORS

Mission Innovation enjoins 20 countries and the European Union[4] to double current annual public R&D funding in clean energy to $30 billion by 2020. (The EU's pledge is based on central research funding and some EU countries have also enrolled separately.) Global Apollo, meanwhile, proposed investing $15 billion a year for ten years[5]. It calls on developed countries to plough 0.02% of their gross domestic product (GDP) into ▶

public R&D to make electricity from renewable sources cheaper than that from coal by 2025.

The voluntary approach of Mission Innovation can be 'gamed' to lower a nation's commitment. Nations calculate their Mission Innovation pledges by choosing a baseline for funding and doubling it. When countries announced their pledges in June this year, most based them on unreported data or funding statistics that are not clearly defined. Only Australia and Canada used official data published by the International Energy Agency (IEA). Some countries chose a single year (2013, 2015 or 2016) from which to double government spending; others took a three-year average (from 2010 to 2013, say).

Such choices shift the goalposts. For example, Australia's target was 4.5 times lower than it could have been — Aus$208 million (US$160 million) rather than Aus$938 million — because it used 2015 as a starting point (Aus$104 million[4]) rather than using a three-year average from 2012 to 2014 (Aus$469 million). The EU, France, Mexico, Norway, Sweden and the United Kingdom used three-year average baselines (see 'Big promises').

Pledges may partially repackage separately planned spending increases. For example, the EU's clean energy R&D target is €1.974 billion (US$2.2 billion) per year by 2020. It would have reached €1.493 billion by 2020 anyway without Mission Innovation.

A doubling goal continues the global imbalance in R&D capacities. Mission Innovation's national targets as a percentage of GDP vary by a factor of 20, from Chile's 0.0037% to Norway's 0.072%. By contrast, Global Apollo's 'one size fits all' contribution of 0.02% GDP may be hard for developing nations to achieve and too low to make a difference for R&D intensive countries.

WHAT DOES CLEAN MEAN?

The scope of what is deemed clean energy R&D varies between countries. This makes pledges difficult to decipher. Some speak of just research and development. Some add another 'D': demonstration. Most clean-energy R&D is concentrated in a few areas — the United States, China and Europe. Mission Innovation says little about how to spread advances to other regions and deploy them at scale.

The many definitions of clean energy change budgets dramatically. Some countries take it to mean renewables such as wind, solar and hydropower; others include energy efficiency, nuclear energy and carbon capture and storage (CCS). Some interpret clean energy as that which is non-polluting or has low environmental impact or comparatively few carbon emissions. Some even include 'clean coal' or the increased deployment of natural gas. For example, nuclear power accounts for one-quarter of the United Kingdom's baseline. A country could meet its pledge by tripling nuclear R&D and doing little on electric transport, renewable energy or smart grids, say.

Most countries gave no spending breakdowns by sector, or offer confusing ones. Germany's stated definition of clean energy includes renewable energy, energy efficiency, storage technologies, grid technologies, CCS, fuel cells and other sectors, including 'cleaner fossil energy'. But the country's three-year average baseline for Mission Innovation (annual R&D funding between 2013 and 2015) of €450 million seems only to reflect expenditure on renewable and energy-efficient technologies (€488 million reported to the IEA in 2014) and not that spent on CCS, hydrogen and fuel cells, and power and storage (another €129 million). Without accurate data, it is hard to judge each country's intentions.

These problems weaken progress towards the goals of the Paris agreement. Global advantages from regional research specialities, such as knowledge from Denmark on designs for wind turbines, will be squandered. And Mission Innovation spending will be spread thinly. In June, the partnership published an 'enabling framework' that sets out general principles and ways of working[6]. It lacks detail and concrete next steps.

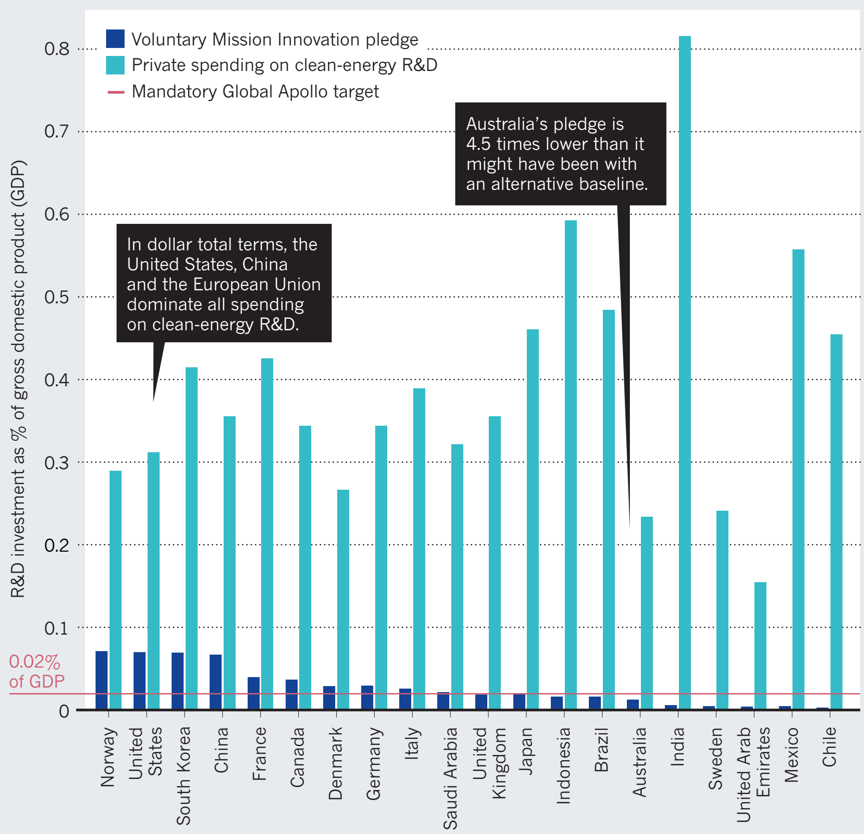
Mission Innovation's leadership should learn from Global Apollo's more directed call for technological change, clear definition of clean energy, transparent investment targets and robust platform for collaboration. Global Apollo focuses on three areas: photovoltaics and concentrating solar power, electricity storage and smart grids[5]. It has one goal: plugging a steady supply of low-cost renewables into the grid. Mission Innovation will match Global Apollo's investment — $150 billion over ten years. But funds will be spread across many more sectors, including nuclear power (pledged by Australia, Brazil, Canada, China, South Korea, the United Arab Emirates, the United Kingdom and the United States) and industrial energy efficiency (all countries except China)[4].

PRIVATE SECTOR

The private sector dominates R&D in clean energy. It is absent from both Mission Innovation and Global Apollo. Funding levels are hard to establish because much of corporate R&D takes place in-house. It typically

BIG PROMISES

Private spending on clean-energy research and development (R&D) dwarfs the public pledges of nations participating in Mission Innovation and the Global Apollo Programme, two initiatives intended to boost national clean-energy R&D.



SOURCE: SEE SUPPLEMENTARY INFORMATION; GO.NATURE.COM/2CDCNQK

XHANGW HONGXIANG/XINHUA/EYEVINE

Solar panels in northwest China's Qinghai province.

accounts for a large proportion of a country's R&D — around 70% of all R&D in the United Kingdom[7], for instance. The overall proportion may be even higher in other nations, such as India and Chile.

Private-sector spending on R&D can be estimated by tracking chains of transactions between companies and reported figures of proportional spend on R&D (transactional data). This methodology underlies the 'Low Carbon and Environmental Goods & Services' data set developed by the digital intelligence company kMatrix (of which M.P. is director)[8]. Our findings suggest that previous, general assessments underestimate private funding for clean energy R&D.

Private investment is mainly directed at technologies deployed at scale rather than those in development. For example, in the United States, we found that for every dollar of public R&D funding reported to the IEA, private companies invest \$25 in renewables R&D but just \$0.56 in CCS. Similar analyses would help countries to identify other areas that are not being backed by private companies — perhaps hydrogen and fuel cells — and thus need more public support.

Partnerships must be forged. Public R&D is not just 'blue sky' exploration. It can shape markets and drive innovation in areas where the private sector is risk-averse[9], helping to create markets for new technologies and make technologies viable. And the private sector needs to go beyond its conventional ways of commercializing technologies.

There are positive signs. The Breakthrough Energy Coalition is a group of investors who pledged in Paris last December to support technologies arising from Mission Innovation with 'patient capital'. They will make important, long-term investments, instead of backing companies for the quickest profit.

PLEDGE ALLEGIANCE

For Mission Innovation to revolutionize our global energy system, more governments must sign up and all countries must meet their 2020 pledges. Some members of the Group of 20 industrialized nations and guests (including Argentina, South Africa and Spain) have not yet joined. September's G20 Meeting in Hangzhou, China, was a missed opportunity for more countries to make a high-profile commitment. The Mission Innovation secretariat and steering committee must agree on a mechanism for reviewing and tweaking pledges on the basis of actual research spend using fairer baselines and a sensible, shared definition of clean-energy innovation.

"Countries must meet their 2020 pledges."

Governments need to fund both research into radical new technologies and targeted development with commercial potential. Mission Innovation can use its political goodwill to ensure that countries work closely together to share new clean technology and deploy it at a global scale. Such a change can be achieved only if member countries voluntarily put close collaboration before national priorities.

As economist Mariana Mazzucato put it[9], there needs to be a symbiotic relationship — rather than a parasitic one — between state-funded R&D and the private sector. Public innovation funding needs to do three things better: set priorities for private R&D; drive greater collaboration between state-funded early-stage research and privately funded translation; and incentivize the private sector to bring new technologies to market.

We urge governments to use studies of transactions, as illustrated here, to examine what private R&D offers to the clean-energy equation and direct the extra funds from their pledges into areas that are currently underdeveloped. ■

Lucien Georgeson *is a doctoral researcher, and* **Mark Maslin** *is professor, in the Department of Geography, University College London, UK.* **Martyn Poessinouw** *is director of kMatrix Ltd, Greetham, UK.*
e-mail: lucien.georgeson.13@ucl.ac.uk

1. Hare, B. *et al. Policy Brief: Below 2°C or 1.5°C depends on rapid action from both Annex I and Non-Annex I countries* (Climate Action Tracker, 2014).
2. Sánchez, J. J. *J. Iberian Latin Am. Econ. Hist.* **27,** 141–174 (2010).
3. Allen, T. & Arkolakis, C. *Q. J. Econ.* **129,** 1085–1140 (2014).
4. Mission Innovation. *Baseline, Doubling, and Narrative Information Submitted by Mission Innovation Countries and the European Union* (2016).
5. King, D. *et al. A Global Apollo Programme to Combat Climate Change* (LSE, 2015).
6. Mission Innovation. *'Enabling Framework' for Mission Innovation* (2016).
7. Economic Insight. *What is the Relationship between Public and Private Investment in Science, Research and Innovation?* (2015).
8. UK Department for Business Innovation & Skills. *Low Carbon and Environmental Goods and Services (LCEGS)* 67 (BIS, 2013).
9. Mazzucato, M. *The Entrepreneurial State* 2nd edn (Anthem, 2015).

Supplementary information accompanies this article online: see go.nature.com/2cdcnqk.

Renewables need a grand-challenge strategy

Launch a global clean-energy initiative to set priorities that galvanize researchers to deliver breakthroughs, write **Alan Bernstein** and colleagues.

Public spending on research into renewable energy is too low to meet even the modest targets set at the Paris climate talks last December, let alone decarbonize the world economy. It stands at about US$6.5 billion a year, or less than 2% of total public research and development (R&D) spending, according to data from the International Energy Agency.

There are encouraging signs that the political will and private-sector interest is coming into place to accelerate the transition to a decarbonized economy. Two recently formed initiatives aim to increase public funding for renewables R&D: Mission Innovation and the Global Apollo Programme (see page 27). In addition, the Breakthrough Energy Coalition is a group of private-sector investors (led by Bill Gates) who have pledged to invest in innovative ideas resulting from publicly funded research.

To ensure the strategic and most effective use of these funds, stakeholders must work together across countries, sectors and disciplines. We propose that this is done through a 'grand challenges' strategy for renewable energy. We argue that agreeing on a global set of priorities would be an efficient and effective way for countries and funders to make decisions about which technologies to back.

As we approach the first anniversary of the Paris climate agreement, we urge governments, researchers, the private sector and philanthropists to act quickly to forge this important initiative.

TO-DO LIST

Grand challenges are proven to accelerate scientifically risky research in areas that would otherwise be left behind, such as global health.

In 2005, the Bill & Melinda Gates Foundation in Seattle, Washington — with the Wellcome Trust in London, the Canadian Institutes of Health Research and the Foundation for the US National Institutes of Health — targeted issues surrounding neglected diseases, which affect most of the world's population (H. Varmus *et al. Science* **302,** 398–399; 2003). Fourteen priority topics included developing a genetic strategy to incapacitate insects that transmit agents of disease, such as the mosquito vectors of yellow fever, dengue and Zika virus (D. A. Joubert *et al. PLoS Pathog.* **12,** e1005434; 2016).

Renewable energy calls for a broadly similar approach. It is a difficult, urgent global problem that has been neglected in terms of public research and investment. It requires big thinking, multidisciplinary approaches and supportive policies to compete with existing systems. And it is tightly coupled to other global challenges, such as food and water security, poverty and health.

> ***"Scalability, affordability, uptake and dissemination need to be addressed."***

We propose the following steps. A consortium of funding partners, including some or all of the Mission Innovation countries, Breakthrough Energy Coalition investors, philanthropic foundations and other private-sector actors, should appoint an international science board of distinguished researchers, policymakers, captains of industry and engaged citizens from developed and developing countries.

The board's task would be to distil a set of grand challenges for renewable energy, and make them as detailed as those for global health. Areas to be addressed include energy harvesting and storage, smart grids and transmission, policy levers and economic models.

Take energy storage, for example. One challenge might be to produce large- and small-scale storage systems that are safe, scalable and inexpensive. This would catalyse a range of research, from batteries based on simple 'flow' systems — which host a series of reactions in one device and use Earth-abundant materials for electrolytes and membranes — to new science and engineering for storing energy as liquid fuels. Any storage system must be safe and sustainable, competitive and compatible with energy generation and distribution systems.

SHARED PURPOSE

Scalability, affordability, uptake and dissemination need to be addressed, and links improved between science and policy. In the developed world, public policies — including feed-in tariffs — might be required to encourage the development of disruptive innovations that hold promise to displace existing technologies. In the developing world, new technologies must respond to basic local needs such as food and water security. Near-term strategies for improving energy efficiency and reducing carbon release must be delivered with long-term ones to decarbonize the global economy.

Such a shared purpose would align the efforts of governments, agencies, foundations, investors, the research and policy communities, citizens and industry. It would accommodate the many disciplines needed across the natural and social sciences, and galvanize the best investigators — regardless of country — to work together to help solve one of the world's most pressing problems. ■

Alan Bernstein *is president and chief executive of the Canadian Institute for Advanced Research (CIFAR), Toronto, Canada.* **Edward H. Sargent** *is a CIFAR senior fellow and director of CIFAR's Bio-Inspired Solar Energy Program in the Edward S. Rogers Sr. Department of Electrical and Computer Engineering, University of Toronto, Canada.* **Alán Aspuru-Guzik** *is a CIFAR senior fellow, and professor of chemistry and chemical biology at Harvard University, Cambridge, Massachusetts, USA.* **Richard Cogdell** *is a CIFAR adviser, and professor of botany at the University of Glasgow, Glasgow, UK.* **Graham R. Fleming** *is distinguished professor of chemistry at the Kavli Energy NanoSciences Institute, University of California, Berkeley; and at Lawrence Berkeley National Laboratory, Berkeley, USA.* **Rienk Van Grondelle** *is a CIFAR senior fellow, and professor of biophysics at Vrije Universiteit Amsterdam, the Netherlands.* **Mario Molina** *is professor in the Department of Chemistry and Biochemistry, University of California, San Diego, and president of the Mario Molina Centre for Strategic Studies on Energy and the Environment, Mexico.*
e-mail: alan.bernstein@cifar.ca

A list of co-signatories accompanies this article online: see go.nature.com/2dtppt9.

AUTUMN BOOKS

THEORETICAL PHYSICS

Windows on the weird

Robert P. Crease weighs up a theoretical-physics study that cracks open a strange vista.

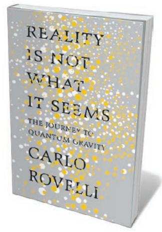
Can you explain loop quantum gravity to people who know next to nothing about physics? Carlo Rovelli's *Reality Is Not What It Seems* shows that you can. Following the physicist's acclaimed *Seven Brief Lessons on Physics* (Allen Lane, 2015; R. P. Crease *Nature* **526,** 37–38; 2015), this book invites the reader to see "through the window" into the beautiful and surprising world of contemporary theoretical physics. Its only drawback is an annoying and unnecessary presumption, announced by the title, that suggests that the view out of the window is into reality itself.

In most respects, this book is a model of popular science writing. The first half provides a select series of vistas on early thinkers and their ideas, to prepare the ground for the adventure of the second half. Rovelli evokes classical Greek philosophers Aristotle and Democritus to familiarize us with the question of whether, at base, the natural world is continuous and smoothly varying, like a beach seen from afar, or grainy, like a beach seen close up — that is, with no arbitrarily small amount of matter. Isaac Newton's seventeenth-century vista shows us reality as an infinite space in which time passes and particles push each other about with mathematically describable forces, such as his laws of motion and universal gravitation. Michael Faraday and James Clerk Maxwell contribute two new things:

ILLUSTRATIONS BY EOIN RYAN

electromagnetism, which fuses electricity and magnetism into a single thing, and the idea of a field, or something suffused throughout space that acts and is acted on by electric and magnetic particles.

In the twentieth century, the entire landscape changes. Quantum mechanics fuses particles and fields, makes them indeterminate and implies that things exist only when interacting with other things. Einstein fuses space and time into space-time, then treats Newton's space as nothing more than the gravitational field itself. This effectively 'curves' space, making the world finite but without boundaries. Such a space can be described either 'from without' as a mathematical representation, or 'from within', as what a putative person on horseback, say, would encounter when travelling through it. It turns out, Rovelli shows, that such a rider, journeying in a straight line, would end up back at the point of departure, thereby traversing a loop. That provides a key image for what is to follow.



Reality Is Not What It Seems
CARLO ROVELLI
Allen Lane: 2016.

But the views through the relativity and quantum windows differ. The world of the former is curved, precise and continuous; of the latter, Euclidean, indeterminate and discrete, with no arbitrarily small amounts of matter or energy. This tension outlines the problem of twenty-first-century theoretical physics. Numerous popular-science books have covered the journey thus far, but here — about halfway through — Rovelli's becomes unique. From this point on, he aims to provide "live coverage of the ongoing research" on the particular strategy that he and his colleagues are adopting in the ambitious quest to unite relativity and quantum mechanics.

The quest was launched by the Wheeler–DeWitt equation, which describes space at small scales as having something like the frothiness of quantum fields, with no arbitrarily small amounts of space. The economy and care with which Rovelli has prepared the reader now pays off, as he uses the vistas presented in the first half of the book to assemble a portrait of loop quantum gravity. Faraday and Maxwell's description of electric force in terms of lines that close or loop around resembles how the Wheeler–DeWitt equation describes gravitation. But whereas electromagnetic lines are infinitely fine and continuous, the gravitational lines of the Wheeler–DeWitt equation are quantized — patchy and distinct like a spiderweb. Finally, Einstein used gravitation to structure space in a similar way to how the Wheeler–DeWitt equation uses loops. In a nutshell: loop quantum gravity is Faraday's lines plus the granularity of quantum theory plus Einstein's idea that these lines are the structure of space. For good measure, Rovelli includes a picture of a T-shirt emblazoned with the basic equation of quantum loop gravity.

Other quests have the same goal. String theorists, for instance, have embarked on a different path to unite gravitation and quantum mechanics. In favouring quantum loop gravity, Rovelli is conservative: he is not relying on radically new ideas. Yet his approach has radical consequences, on which he spends the rest of the book elaborating. Space is granular (with no arbitrarily small volume); time does not exist (there's no variable for it in the Wheeler–DeWitt equation); and the basic stuff of the world consists of special kinds of quantum field. What we see through the window is utterly unlike our conventional world. No problem! This, he writes, is simply because we humans are like moles living underground to whom someone describes the Himalayas. Or like the people in Plato's cave, whose imaginations are chained by prejudice, ignorance and our senses, and who can view only shadowy representations of the real. Rovelli confidently puts reality on the other side of the window from the "parochial experience" in which we ordinary humans live and work.

SPACE IS GRANULAR; TIME DOES NOT EXIST; AND THE BASIC STUFF OF THE WORLD CONSISTS OF SPECIAL KINDS OF QUANTUM FIELD.

A sceptic might react to this irksome scientism by objecting that, unlike in Plato's image, the vistas seen through the window keep changing. One can imagine, too, a book by a string theorist offering another view out of the window — just how many exits does Plato's cave have? Yet another problem is that Rovelli has a cavalier attitude towards philosophy. Plato's cave is more nuanced than he makes out, and Rovelli misinterprets a passage to claim that Socrates was disappointed by scientists. He plucks a statement out of context from a lengthy autobiographical story in which Socrates is describing youthful views from which he has since moved on (Rovelli also misquotes the translation).

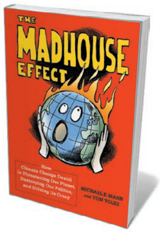
Philosophers tend to have a more existential take on 'reality', not restricting it to what scientists represent but seeing it as also encompassing something of what the moles and horsemen encounter. The regrettable thing is that the scientism in this otherwise fine work is unnecessary, although I know it helps to sell books and cement the prestige of science. That's just today's reality. ■

Robert P. Crease *is a professor in the Department of Philosophy at Stony Brook University, New York.*
e-mail: robert.crease@stonybrook.edu

CLIMATE SCIENCE

Denialism deciphered

Dave Reay enjoys a wry history of US climate-science obfuscation.



The Madhouse Effect: How Climate Change Denial is Threatening Our Planet, Destroying Our Politics, and Driving Us Crazy
MICHAEL E. MANN AND TOM TOLES
Columbia University Press: 2016.

As an iconic climate-change image, the 'hockey-stick graph' by geophysicist Michael Mann — showing global temperature change over the past 1,000 years — is up there with the greats. Others include the Keeling curve of changing atmospheric carbon-dioxide concentrations and the 'boiling frog' metaphor from Al Gore's 2006 documentary *An Inconvenient Truth*. Mann's figure (from a seminal paper: M. E. Mann *et al. Geophys. Res. Lett.* **26,** 759–762; 1999) appears in 'Climate Science 101' lectures the world over; was a touchstone of the 2001 third assessment report of the Intergovernmental Panel on Climate Change; and still elicits invective from deniers (S. Lewis *Nature* **483,** 402–403; 2012). Who better than Mann, then, to explore the history of climate-change denial, and its politics, personalities and implications?

The Madhouse Effect is a breezy, engaging read, interspersed with wry illustrations courtesy of cartoonist Tom Toles of *The Washington Post*. It offers many excellent insights into life on the front line battling US climate-science obfuscation. We learn about the cadre of contrarian scientists routinely rolled out to cast doubt on issues such as ozone depletion and anthropogenic climate change (as well as second-hand smoke and the dangers of pesticides). We read of the television, radio and Internet 'shock jocks' who chase ratings by giving equal weight to scientific consensus and denialist rhetoric. The power of vested interests in US politics and implications for state and federal action on climate change are made abundantly clear, with Mann an amiable, if rather despairing, guide.

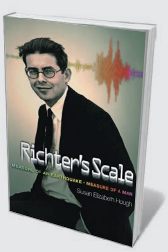
He begins with an overview of the scientific method, the science of global warming and key uncertainties — such as feedback mechanisms, whereby warming can itself boost greenhouse-gas emissions and so cause even more warming. He and Toles then explore the "six stages of denial", ranging from 'it's not happening' through 'it's self-correcting' to 'geoengineering will fix it all'.

Where this book shines is in its exploration of the debate in the United States, and a veritable who's who of denial. As the November presidential election looms, it's useful to learn about key players' stances. Unsurprisingly, most of the contenders for the Republican nomination when the book was finished back in July emerge as outspoken critics of climate science and international action. The party's current candidate, Donald Trump, wants to renegotiate or leave the 2015 Paris climate agreement joined by President Barack Obama in September, and has called climate change a hoax. But Mann suggests that several candidates were influenced by cryptic political and financial forces in the fossil-fuel industry, which apparently bankroll denialist activity and lobbying to protect their interests.

The authors discuss how Republican senator Jim Inhofe (Oklahoma) is waging a "war" on climate science by using hearings of the Senate environment committee that he chairs to try and debunk climate change. Mann's writing is subjective in places — such as when discussing former Virginia attorney-general

NEW IN PAPERBACK

Highlights of this season's releases



Richter's Scale: Measure of an Earthquake, Measure of a Man
Susan Hough (Princeton Univ. Press, 2016)
Charles Richter's eponymous, logarithmic scale of earthquake classification made him globally famous. In this illuminating biography, seismologist Susan Hough describes Richter's accidental arrival at the Seismolab of the California Institute of Technology, and the colleagues there who resented his fame. A surprising selection of Richter's poetry surfaces, reflecting his sentiments on married life and mortality (see Gregory Beroza's review: *Nature* **445,** 599; 2007).

Ken Cuccinelli, an erstwhile alleger of data manipulation, now an oyster farmer on an island threatened by rising sea levels. But he generally manages to avoid score-settling.

In 2009, Mann's work was caught up in the 'Climategate' scandal (nature.com/climategate). This was the unauthorized release of more than 1,000 e-mails from the Climatic Research Unit at the University of East Anglia in Norwich, UK — many containing private correspondence, some to or from Mann. Excerpts were published by climate sceptics to smear scientists and cloud public and political judgement. Mann gives this seismic event just a couple of pages. He explains briefly how the e-mails were taken out of context and that references to a "trick" used to "hide the decline" referred simply to a trick of the trade: combining direct measurements of global temperature with proxy estimates. Given that Mann was bombarded with threats and abuse following Climategate, a fuller exploration — as in Fred Pearce's *The Climate Files* (Guardian Books, 2010) — would have been good to see.

Despite the political tensions, Mann and Toles strike a positive tone in the final section. They highlight action being taken at community, city and state levels, and the potential of the Paris agreement to avoid the most damaging effects of climate change. And they find hope in the power of individual choice to shift the most recalcitrant hangovers from our carbon-intensive history. Their key recommendations are for each of us to support renewable energy and carbon pricing, to vote for politicians who do the same and to stop equivocating on climate science.

As Mann points out, denialists are not likely to read this book. For climate researchers outside the United States, it is an eye-opening primer (despite its baffling references to baseball stars) on the vested interests with which their US colleagues must do battle. For a wider readership, it makes clear just how high the stakes are. If tackling climate change is indeed a war, then Mann and Toles have certainly earned their stripes. I salute them. ■

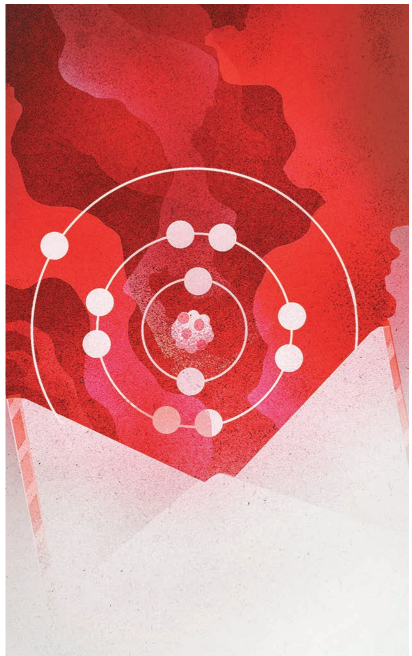
Dave Reay *is chair in carbon management and assistant principal for global environment & society at the University of Edinburgh, UK, and author of* Nitrogen and Climate Change.
e-mail: david.reay@ed.ac.uk

MEMOIR

Love and uncertainty

Werner Heisenberg's wartime letters to his wife record scientific and personal privations, finds **Ann Finkbeiner**.

My Dear Li: Correspondence, 1937–1946
WERNER HEISENBERG AND ELISABETH HEISENBERG; ED. ANNA MARIA HIRSCH-HEISENBERG, TRANSL. IRENE HEISENBERG
Yale University Press: 2016.



Werner Heisenberg is a conundrum. He won the 1932 Nobel Prize in Physics for creating the foundations of quantum mechanics and his uncertainty principle, which describes how it is impossible to know a particle's location and its momentum simultaneously. During the Second World War, directed by the Nazi government, he headed Germany's unsuccessful efforts to create an atomic bomb. Why didn't he succeed? Why did he try?

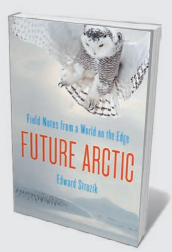
There are no unambiguous answers here, although clarifying Heisenberg's motives is one reason that his daughter, Anna Maria Hirsch-Heisenberg, gives for publishing the letters between him and her mother. What the letters do illustrate is Hirsch-Heisenberg's other reason for publishing (in German in 2011, and now in English for the first time): how a couple much in love lives through a war.

Werner begins his letters with "My dear Li". Li is Elisabeth, née Schumacher; they met in 1937 at a musical evening. The two talked — a conversation, Werner wrote, that seemed to have begun so long ago that continuing it for the rest of their lives felt natural. Two weeks later, they were engaged; four months later, they began a 40-year marriage. But Heisenberg had to travel for research and was rarely at home, thus the letters. This collection spans the tumultuous years from 1937 to 1946.

The letters, necessarily discreet about politics and the military, contain mostly the quotidian — frighteningly so, this being Germany during that war. By 1939, Werner lives in Leipzig and Li has moved to their safe country house in southern Germany. Li has had twins; she will have four more children in the next five years. The war has started. "I get caught up pondering the dark picture everybody is painting," Li writes, "how fortunate that the children ... are so unencumbered and jolly." Werner makes a long lecture trip to the United States, where he finds the audiences receptive and the students bright. He tells his US colleagues who offer him jobs that he needs to stay in Germany "so that I might also be here afterward and help"; as he writes to Li, "we are just not at home here".

Over the next few years, Werner alternates between Berlin, where "it is quite ▶

How to Clone a Mammoth
Beth Shapiro (Princeton Univ. Press, 2016)
Ecologist Beth Shapiro parses possible impacts of the "unextinct". Reintroducing mammoths to Siberia, for example, could restore grasslands and keep carbon trapped in the permafrost (see Henry Nicholl's review: *Nature* **521,** 30–31; 2015).



Future Arctic: Field Notes from a World on the Edge
Edward Struzik (Island, 2016)
Arctic journalist Edward Struzik compresses 30 years of circumpolar observation in this portrait of a thawing world. As warmer oceans induce powerful storms that hasten the ice's retreat, ecological anomalies surface, such as the grizzly bear–polar bear hybrid.

striking these days how everybody becomes thinner", and Leipzig, where newspapers carry obituaries of young people dying. "I myself am often so sad and downcast," he writes to Li, "without you I would not quite be able to cope". Food is scarce; Werner preserves cherries from his Berlin garden. His work, directing research on nuclear fission, "makes no sense".

In 1945, between air raids, Werner advises Li that as the front moves closer to southern Germany, she should watch for attack planes and the children should practise throwing themselves to the ground near a wall. Li makes her own yeast and worries about getting enough flour for bread. They tell each other that they are thinner and more exhausted. "Love," he writes, "stay well and prepare for the more difficult times."

Near the war's end, Heisenberg and other German nuclear scientists are arrested by the Allies. They are held for six months in England; few letters are allowed. For lack of food, Li puts two of the children into a home. She cares for Heisenberg's dying mother and cuts their firewood. He's released in January 1946. "I want to build a containing wall around you from all the love I have in my heart," writes Li. The letters end that June, with the family reunited and living in Göttingen; in 1950, they have a seventh child.

Hirsch-Heisenberg writes that the letters were chosen and edited for relevance and concision. We cannot know what other filters, if any, children apply to the publication of their parents' letters. Hirsch-Heisenberg gives no sources, but makes the case that her father's motives for working on a German atomic bomb were to control atomic research and to convert it to peaceful uses, but that building an actual bomb was "out of the question". Judging from these letters, Heisenberg was doing what it took to wait out the dreadful storm so that he could get on with his life with physics and Li. ■

Ann Finkbeiner *is a science writer in Baltimore, Maryland.*
e-mail: anniekf@gmail.com

THEORETICAL PHYSICS

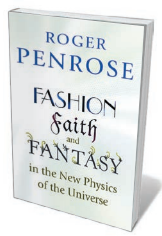
The emperor's new physics

Richard Dawid examines a critique of quantum mechanics, string theory and inflationary cosmology.

The eminent theoretical physicist Roger Penrose is worried about the current path of physical research. In *Fashion, Faith, and Fantasy in the New Physics of the Universe*, he argues that the eponymous triad of trends has become overly powerful in contemporary fundamental physics. This core message is delivered in language that demands some mathematical sophistication of the reader. Penrose also discusses some of his own ideas, such as twistor theory — his take on a synthesis of quantum theory and general relativity.

Penrose claims that even well-confirmed theories, such as quantum mechanics, are 'oversold' with respect to their presumptive stability. Quantum physics has had an impressive record of predictive success, ranging from quantum chemistry to elementary particle physics. But it faces a deep conceptual problem. Whereas quantum mechanics has a perfect internal consistency when it describes a system that evolves without being measured, the way in which it represents measurements is not coherently embedded in that description. To Penrose, this indicates that the fundamental principles of quantum mechanics have not yet been found and will rely on the elusive full integration of gravity into quantum physics. He argues that the success of quantum mechanics tends to make physicists insensitive to the theory's conceptual problem and generates an unjustified degree of faith in its basic principles as a solid foundation of physics.

Another source of undue trust in a theory, Penrose asserts, is the physics community's tendency to follow fashion — that is, to settle on one strategy of dealing with a problem before severely testing the theory's empirical predictions. Penrose views string theory (a theory of quantum gravity) as the pre-eminent example.

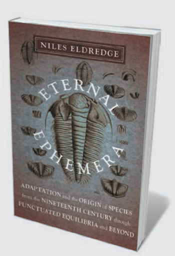


Fashion, Faith, and Fantasy in the New Physics of the Universe
ROGER PENROSE
Princeton University Press: 2016.

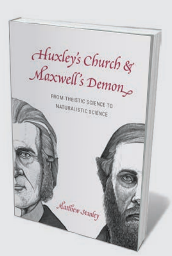
The final trend in Penrose's triad is fantasy — that is, a wildly speculative idea that goes far beyond what is implied by the known data. Penrose assigns that category to inflationary cosmology, which he argues is treated as an established theory despite a lack of evidence.

Of these three, Penrose's discussion of quantum mechanics ('faith') is the most successful. On the basis of an inspired presentation of quantum mechanics, he makes a case that the theory's enormous scientific success does not remove serious doubts about the finality of its basic principles. His discussions of fantasy and fashion, however, are problematic. He paints an exaggerated picture of their role and systematically underrates the merits of the theories he criticizes.

Fashion and fantasy are presented in separate chapters as independent influences that have become too powerful. But, as Penrose acknowledges, fantasy has always been at the root of new theories. Just think about the atomist speculations that led to the kinetic gas theory in the nineteenth century. For Penrose, the trouble arises when fantasy



Eternal Ephemera
Niles Eldredge (Columbia Univ. Press, 2016)
Palaeontologist Niles Eldredge presents an insightful history of evolutionary biology, from transmutation's forefather, Jean-Baptiste Lamarck, comparing fossil molluscs in 1801, to the theory of punctuated equilibria, whereby rapid speciation disrupts periods of stasis.



Huxley's Church & Maxwell's Demon
Matthew Stanley (Univ. Chicago Press, 2016)
The context of Victorian science swung smoothly from the theistic to the naturalistic, shedding supernatural causality along the way. Matthew Stanley attributes the relative amity between Christian and atheist scientists to shared ideals such as intellectual freedom.

is given too much credit before a theory is empirically tested. This occurs, he says, when a theory becomes the subject of fashion. In this light, it is difficult to see the independent role of 'fantasy' in Penrose's argument.

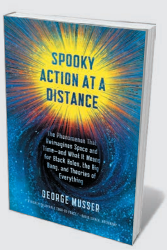
Inflationary cosmology is, moreover, not a good illustration of fantasy, even by Penrose's own account. As he acknowledges, recent precision measurements of the cosmic microwave background agree with typical predictions of inflationary cosmology, so it seems difficult now to call it a mere flight of fancy. Penrose presents his important criticism that inflation generically does not explain the low initial entropy of the Universe (although explanations have been suggested in certain models; see S. M. Carroll and J. Chen. https://arxiv.org/abs/hep-th/0410270; 2004). But he presents the case against inflation in a way that hides the independent significance of problems that can be solved by it, such as explaining the homogeneity and flatness of the observed Universe.

There are similar issues with Penrose's claim that fashion is the main reason for string theory's influential position. His analysis of its problems is not up to the task of debunking proponents' physics-based reasons for confidence. Penrose's main complaint about string theory is that it lacks a clear specification of its number of degrees of freedom. He tries to show this in several contexts. However, he tends to omit information that could make the situation less confusing than he takes it to be. For example, he expresses unease about 'gauge–gravity duality', the claim that string theory is empirically equivalent to a quantum field theory in a lower-dimensional space. (If generally valid, that would mean that a string theory in three extended spatial dimensions was empirically equivalent to a quantum field theory in two spatial dimensions.) Such a claim looks startling, because one would naively expect that a three-dimensional theory has more degrees of freedom than a two-dimensional one. Penrose presents this as one of many questionable implications of string theory.

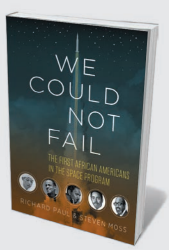
Curiously, however, he presents his case without mentioning that Gerard 't Hooft, who is cited in the book, provided a general understanding of the reduced number of degrees of freedom in quantum gravity without any reference to string theory, before cases of gauge–gravity duality were conjectured in the context of string theory (G. 't Hooft. https://arxiv.org/abs/gr-qc/9310026; 1993). In this light, by generating examples of gauge–gravity duality, string theory does not, as Penrose maintains, make one more prima facie implausible claim, but opens up perspectives for a more thorough understanding of a characteristic of quantum gravity that had already been suggested.

It is always inspiring to read Penrose's uncompromisingly independent perspective on physics. He seems more at home with developing visionary ideas than with detailed criticism of prevalent theories. Unfortunately, this book offers too few of the former and too much of the latter. ■

Richard Dawid *is a philosopher of science at the University of Stockholm.*
richard.dawid@philosophy.su.se



Spooky Action at a Distance
George Musser (Scientific American/Farrar, Straus and Giroux, 2016)
Bending time, space and minds, George Musser investigates nonlocality — two distant particles acting in harmony. With lessons in photon entanglement, particle teleportation and string theory, he ponders how space evolved after the Big Bang.



We Could Not Fail: The First African Americans in the Space Program
Richard Paul and Steven Moss (Texas Univ. Press, 2016)
Profiling NASA's first ten black employees, Richard Paul and Steven Moss show what the space age meant for African Americans. In 1962, NASA granted US$181,000 to a study of the space programme's impact on race relations.

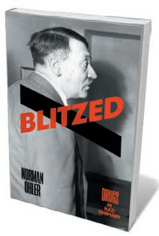
HISTORY

Addiction and the Reich

Paul Weindling ponders a study of drug use among the Nazi leadership and military.

Norman Ohler's *Blitzed* depicts the pervasive drug culture that allegedly developed in Germany's Third Reich. From 1933 to 1945, Adolf Hitler, many Nazi officials and a proportion of the military rank and file were — he contends — in thrall to prescription and recreational drugs. Ohler's is a vivid account; whether it convinces is less certain.

Historians now recognize that despite Nazi racial and political persecution of German scientists, Hitler's Reich offered immense opportunities to many. There was an upswing of research in pharmacology during preparation for total war, and new drugs were hailed as additions to the armoury of high-performance medicine. However, drugs were viewed paradoxically in the Reich. The Nazi ideology of fitness meant that users of opiates such as morphine were branded 'psychopathic personalities', and serious addicts could be compulsorily sterilized. Yet Nazi officials took high-performance drugs such as methamphetamine hydrochloride (crystal meth) and cocaine. German military units and aviators were dosed with the patent methamphetamine-based drug Pervitin (manufactured in Germany from 1937) to improve operational efficiency. And drugs such as Pervitin and metabolic stimulants were tried out on students, military recruits and, eventually, in concentration camps. Questions remain, however, over precisely how the drugs were tested, prescribed, distributed and used.



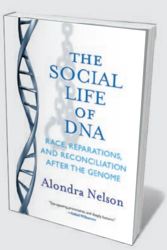
Blitzed: Drugs in Nazi Germany
NORMAN OHLER
Allen Lane: 2016.

Meanwhile, Hitler's façade was that of a vegetarian and non-smoker, but he became increasingly dependent on patent vitamin tonics produced from bovine thyroid glands, livers and bones. (His favoured physician, Theodor Morell, claimed exclusive rights to process these in occupied Ukraine.) Hitler also had a predilection for sedatives such as the opiate Eukodal (oxycodone). But the extent to which he took any of these drugs remains controversial.

Ohler has effectively written two separate books, one focusing on the military; the other on Hitler and Morell. There is a thin connecting thread attributing Hitler's military misjudgements to drugs, such as Operation Barbarossa against the Soviet Union when he overstretched German troops. Ohler's descriptions of military operations, such as the Blitzkrieg against Poland and France, are very generalized. By contrast, the treatment of Hitler as a patient is a detailed study based on Morell's diary and personal records.

Called 'Patient A' by Morell, Hitler is depicted as increasingly stooped and tremor-ridden. Ohler attributes this decline to drug dependency rather than Parkinson's disease (mooted as early as 1945 by the Nazi neurologist Max de Crinis). And Ohler goes further. He argues that drug consumption initially boosted the Reich's military success, but then undermined it as widespread addiction set in — a focus that ignores corrosive factors such as anti-Semitism, the Holocaust and the drive to secure Lebensraum (territorial living space). Ohler presents some staggering statistics on drugs supplied to individual units, but fails to provide statistics on Pervitin production, fluctuations in its supply to the military, or the extent and duration of its use. Nor is there a detailed analysis of individual soldiers to determine the impact on health. If military operations were so saturated by drugs, more evidence should be forthcoming.

Ohler draws on the same sources as other books and papers on the Nazi consumption of crystal meth and cocaine, and on Hitler's medical predilections. These include Giles Milton's *When Hitler Took Cocaine and Lenin Lost His Brain* (Picador, 2016) and the paper 'Speed in the Third Reich' (S. Snelders and T. Pieters *Soc. Hist. Med.* **24,** 686–699; 2011). Some material countering Ohler's argument is not referenced. In *Was Hitler Ill?* (Polity, 2012), Henrik Eberle and Hans-Joachim Neumann argue that Morell was a competent diagnostician, albeit compliant in prescribing. And Ohler does not draw on witness studies such as that of SS nutritionist Ernst-Günther Schenck on Hitler as a patient. Finally, Pervitin was known by the



The Social Life of DNA: Race, Reparations, and Reconciliation After the Genome
Alondra Nelson (Beacon, 2016)
Geneticist Alondra Nelson analyses the rise in DNA 'roots' testing among African Americans seeking their lost identity. Race, politics and science emerge as intertwined as the double helix itself (see Fatimah Jackson's review: *Nature* **529,** 279–280; 2016).

Evolving Ourselves: How Unnatural Selection and Nonrandom Mutation Are Changing Life on Earth
Juan Enriquez and Steve Gullans (Current, 2016)
In this study of the evolution of evolution, Juan Enriquez and Steve Gullans ponder the potential of genome editing and synthetic life. Could pig lungs, 'humanized' by the addition of our genes, obviate human transplants?

1980s to be a crucial component of Nazi high-performance medicine.

Ohler pays more attention to the perpetrators of Nazi drug experiments than to their victims. He cites experiments with mescaline, trying to create a 'truth' drug, from the perspective of the Dachau doctor Kurt Plötner, for instance. My book *Victims and Survivors of Nazi Human Experiments* (Bloomsbury, 2014) draws on original findings from more than 15,000 prisoners' narratives of coerced medical testing — including of mescaline — at Auschwitz and Dachau, but is not referenced. Nor does Ohler mention the victims of another notorious experiment. Seven British Royal Navy commandos endured experimentation with stimulants, including cocaine and amphetamines, at the Sachsenhausen concentration camp. After a forced, three-day march carrying heavy loads, five were executed in 1945. Ohler mentions only a German survivor.

But my key issue is with Ohler's central claim that Pervitin and Eukodal induced a sense of invincibility, first enhancing operational boldness, then destroying the Nazis' ability to engage with military collapse. He also concludes that addiction to ever-stronger doses of patent medicines clouded Hitler's judgement on strategic issues concerning Dunkirk and Crimea. He reduces every twist and turn of the war on the German side to addiction. Yet the US and UK military used amphetamines as part of a highly successful scientific and technological war effort without apparent issues with addiction.

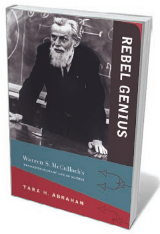
Ohler ends at what he dubs the "Last Exit Bunker", with Hitler addicted to Eukodal. That title encapsulates my problem with *Blitzed*. It strings the reader along with facile phrases such as "High Hitler" and "One Reich, One Dealer", calling the bunker-bound Führer a "super-junkie". This is a text full of short cuts and speculation rather than a balanced synthesis of a mass of literature and sources to date, rendered readable and accessible. ■

Paul Weindling *is research professor in the history of medicine at Oxford Brookes University, UK.*
e-mail: pjweindling@brookes.ac.uk

CYBERNETICS

A mathematician of mind

Manuel Blum examines a biography of cybernetics pioneer Warren McCulloch and his revolutionary times.



Rebel Genius: Warren S. McCulloch's Transdisciplinary Life in Science
TARA ABRAHAM
MIT Press: 2016.

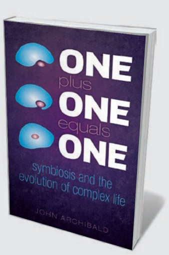
In 1958, in my junior year at the Massachusetts Institute of Technology, Richard Schoenwald — whose tutorial on Sigmund Freud I was taking — encouraged me to meet the anti-Freud, Warren McCulloch. Where Freud had written *The Future of an Illusion* (1927), a critique of religion, McCulloch countered with *The Past of a Delusion* (1953), a reference to Freud (the title says it all). I dropped into McCulloch's basement lab and found myself facing a tall, striking character: long beard, coarse Scottish wool suit, books piled to the ceiling. I confided that I wanted to understand how the brain works. He handed me a sheaf of his 'Research Laboratory of Electronics' publications. These showed how to construct neural networks of formal (model) neurons that could control for errors in those neurons. Weeks later, I stated and proved a theorem that his formal neurons could be configured to do what his networks needed. With that, I was in, mentored and inspired by McCulloch for the next six years and counting.

In *Rebel Genius*, science historian Tara Abraham offers a biography of McCulloch (1898–1969) that shines a light on the twentieth-century revolution in the mind sciences and cybernetics — the scientific study of automatic control in animals (including humans) and machines.

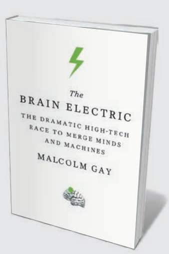
McCulloch insisted that the 'magic' of the brain lay in what electrical networks can do (nowadays, chemistry would count for more). He asserted that the magic would arise whether the networks were constructed from neurons, which he called software (later, meatware) or vacuum tubes, which he called hardware.

> THE 'MAGIC' OF THE BRAIN LAY IN WHAT ELECTRICAL NETWORKS CAN DO.

Like mathematician and computer scientist John von Neumann, McCulloch was interested in errors. Neurons, like vacuum tubes, were unreliable. The problem, he pointed out, was that neuronal thresholds, which affect what neurons compute, are constantly changing. "Thresholds fall when we drink coffee. They rise when we drink alcohol. Yet we can still talk; we can still walk." At least, he could. Computers were then, as now, designed to work with components that make virtually no errors. But at that time, a computer could run for only minutes before errors crept in. How the brain manages with ▶



One Plus One Equals One
John Archibald (Oxford Univ. Press, 2016)
Exuberantly describing the greening of Earth 500 million years ago, John Archibald vivifies the origins of complex life. His microbiologist predecessors star, including Carl Woese, who first sequenced rRNA to track evolution (see Nancy Moran's review: *Nature* **510,** 338–339; 2014).



The Brain Electric
Malcolm Gay (Farrar, Straus and Giroux, 2016)
People enduring amputations, once subject to messy surgery, are now at the forefront of neuroprosthetics research. Malcolm Gay explains the science behind an evolving technology that binds brain impulse to exoskeletons, enabling people with paralysis to move.

faulty neurons was a big question. (And as transistors drop to the size of atoms, errors again become a serious problem in computing.)

McCulloch held sway in a phenomenal period for many fields of science. His multitude of friends and colleagues included neuroscientist Jerry Lettvin, who would drop by to demonstrate one of Hermann von Helmholtz's extraordinary experiments on the eye. Artificial-intelligence pioneer Marvin Minsky showed McCulloch how to construct Venn diagrams of any number of variables (to represent neurons with many inputs). And Manuel Cerrillo convinced McCulloch that he was a genius at filter-design with a self-built hi-fi set that could separate musical instruments from the human voice in a recording.

McCulloch bubbled with ideas. In one co-written paper, 'A Logical Calculus of Ideas Immanent in Nervous Activity' (W. S. McCulloch and W. Pitts *Bull. Math. Biophys.* **5,** 115–133; 1943), he argued that neurons must be capable of inhibition as well as excitation. If not, they would compute only a very small class of 'monotonic' functions. McCulloch told me that neurophysiologists of his time rejected this idea because inhibition had never been observed. His prediction — that inhibition exists in the brain — was later proved experimentally.

Abraham appraises the McCulloch I knew knowledgeably, accurately and insightfully. For example, she writes: "McCulloch's scientific life at its heart was less a philosophical project and much more about transcending disciplines, the power of science to do away with metaphysics, and the power of a neurophysiological, biological psychiatry to eliminate dualist accounts of the mind and non-biological practices in psychiatry." This is both perceptive and accurate.

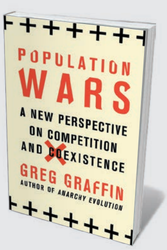
There are also many aspects of McCulloch in Abraham's book that I did not know, a lot that I wanted to know and got, and a lot that I did not even know I wanted to know. For example, Abraham's account of psychologist Clark Hull reveals Hull to be another enormously interesting individual — a proponent of behaviourism who worked in motivation and learning, and who thought that the problem of mind is solvable through scientific theory.

What Abraham does not capture enough of, for my taste, is the striking impression that McCulloch made on his audience — intellectually, through his astute observations, and visually, through his erudite Scottish bearing. Abraham describes a formative experience of McCulloch's: when he was "a student at Haverford College in Pennsylvania in 1917, a teacher asked him what he planned to do with his life". Her version of the event is accurate, but misses the soul of it. What I recall McCulloch saying is that the president of Haverford, Quaker philosopher Rufus Jones, asked, "Warren, what wilt thee be?" to which McCulloch answered, "I don't know." "What wilt thee do?" "I don't know. But," McCulloch added, "I do have a question: 'What is a number that a man may know it, and a man that he may know a number?'" To which Jones rolled back his head and roared, "Thee wilt be busy for the rest of thy life!"

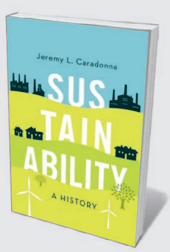
Not everything about McCulloch comes up roses, and Abraham is critical of certain aspects of his approach. She quotes neurophysiologist Ralph Gerard's critique on the Macy conferences on cybernetics — where McCulloch aimed to get psychologists, neurophysiologists, mathematicians and engineers talking. Gerard's words were very much a critique of McCulloch himself. He noted how the group began "in the 'as if' spirit. Everyone was delighted to express any idea that came into his mind, whether it seemed silly or certain or merely a stimulating guess that would affect someone else … Then, rather sharply it seemed to me, we began to talk in an 'is' idiom. We were saying much the same things, but now saying them as if they were so."

McCulloch was a polymath: a neurophysiologist who was also a physician, psychiatrist, poet, writer, architect, engineer and mathematician. His was the all-encompassing intellect that could and did bring these disparate fields together — both in the Macy meetings and in his lab. Through its discussions of McCulloch in the round, *Rebel Genius* is an excellent portrait of the man and his time, and a significant contribution to the history of science. ■

Manuel Blum *is the Bruce Nelson University Professor of Computer Science at Carnegie Mellon University in Pittsburgh, Pennsylvania.*
e-mail: mblum@cs.cmu.edu



Population Wars: A New Perspective on Competition and Coexistence
Greg Graffin (Thomas Dunne, 2016)
Zoologist, geologist and punk rocker Greg Graffin explores how an "us vs them" attitude has infiltrated human consciousness and driven populations to war, despite our unique power to plan our future by reflecting on the past.



Sustainability: A History
Jeremy L. Caradonna (Oxford Univ. Press, 2016)
Historian Jeremy Caradonna chronicles the arc of sustainability from its roots in eighteenth-century European forestry to contemporary local food and zero-waste movements, and its emphasis on balance and the long view over economic growth. Emily Banham

Correspondence

Safeguarding the world's largest lake

Lake Baikal in eastern Siberia is listed as a World Heritage Site by the United Nations because of its exceptional endemic biodiversity. Its ecological and environmental health is now under threat from a government funding cut of almost 30% to the lake's long-term monitoring programme.

Biologists at Irkutsk State University have been sampling water temperature, transparency, and plankton abundance and species composition at weekly intervals, year-round, since 1945. Lake Baikal remained largely pristine in the twentieth century, but its ecosystems are changing fast as surface waters warm and winter ice cover lessens (M. V. Moore *et al. Bioscience* **59,** 405–417; 2009 and S. E. Hampton *et al. Glob. Change Biol.* **14,** 1947–1958; 2008).

In the lake's coastal zone, for example, excessive nutrients from industrial and household pollution are causing mass spread of the green alga *Spirogyra* and die-off of endemic sponges in nearshore waters (O. A. Timoshkin *et al. J. Great Lakes Res.* **42,** 487–497; 2016).

Long-term monitoring of the health of the world's deepest lake is crucial. The cost of sustaining it (less than US$70,000 a year) is vanishingly small relative to the ecological and economic value of this global resource.

Maxim A. Timofeyev* *Irkutsk State University, Russia.*
m.a.timofeyev@gmail.com
**On behalf of 5 correspondents (see go.nature.com/2dr7ghi for full list).*

Centralized pilot for e-waste processing

In Guiyu, China, local government has established an industrial park that concentrates electronic-waste-processing facilities to limit their potential environmental and health impacts (see Z. Wang *et al. Nature* **536,** 23–25; 2016). The park, which is used by some 80,000 people, is also an important source of employment.

Electronic waste in the area was previously manually dismantled in household workshops, with no environmental or health protection. In the processing park, created in 2014, new techniques and specialized facilities remove and protect against pollutants. Volatile pollutants, for example, are collected and piped to a treatment facility. Air sampling and local reports indicate that air quality has significantly improved as a result (unpublished data).

It will take time to fully implement the Basel convention on transboundary waste movement (http://www.basel.int/#2), and even longer for individual countries to formulate strict regulations for the disposal and processing of electronic waste. Meanwhile, the Guiyu model offers a solution for limiting damage to the environment and to public health.

Ya Tang *Sichuan University, Chengdu, China.*
tangya@scu.edu.cn

Open data: policies need policing

Like several other progressive publishers, you now require research papers to include a data-availability statement to ensure that the data are sufficient "to interpret, replicate and build on the findings reported in the paper" (*Nature* **537,** 138; 2016). In my view, compliance should be enforced as a condition of publication.

Examples of laxity by publishers include allowing a data-availability statement indicating that "all relevant data are within the paper", when in fact the article included only summary values, and a quantitative study on open data published — ironically — without archived data in a searchable, online repository (the data set was in the supplementary material, which is not always searchable in subscription journals).

Alarmingly, more than half of the archived data sets in journals that mandate open data are incomplete or deposited in a way that obstructs reuse (D. G. Roche *et al. PLoS Biol.* **13,** e1002295; 2015). The responsibility for enforcing compliance with a data policy is in the hands of a journal's editors and reviewers. This needs to be stated explicitly and resourced adequately. I urge the Nature journals to ensure that the new measures are strong and effective.

Dominique Roche *University of Neuchâtel, Switzerland.*
dominique.roche@unine.ch

Open data: curation is under-resourced

Science funders and researchers need to recognize the time, resources and effort required to curate open data (see *Nature* **537,** 138; 2016). Although organizations such as the US National Science Foundation and the European Commission are aiming to make data repositories financially self-sustaining, this is unlikely to happen within one or two funding cycles.

There is no reliable business model to finance the curation and maintenance of data repositories. Databases therefore often restrict access to subscribers (see, for example, go.nature.com/2dzc59o), curtailing opportunities for interoperability and collaboration.

Curation is not fully automated for most data types. This means that — in the life sciences, for example — many popular databases must resort to time-consuming manual curation to check data quality, reliability, provenance, format and metadata (S. Leonelli *Data-Centric Biology* Chicago Univ. Press; 2016).

Crowdsourcing models are promising in this respect because data producers ensure that the deposited data are accurate and reusable, but these models are still not widely deployed (see go.nature.com/2d6p9kc).

To make open data effective as a research tool, computational and field-specific skills need to mesh. This will ensure that data infrastructures are user-friendly and resilient in the face of vertiginous developments.

Sabina Leonelli *University of Exeter, UK.*
s.leonelli@exeter.ac.uk

Costing recombinant antivenoms

The cost of producing antivenoms from recombinant human antibodies to counter the shortage of animal-derived antisera against snakebites is not as prohibitive as you imply (*Nature* **537,** 26–28; 2016).

We estimate that 500–2,000 kilograms of therapeutically active antibodies would be needed to produce enough antivenom to treat the 1 million or so people bitten annually by snakes in sub-Saharan Africa. On the basis of production data for monoclonal antibodies (N. Hammerschmidt *et al. Biotechnol. J.* **9,** 766–775; 2014) and for oligoclonal antibody mixtures (S. K. Rasmussen *et al. Arch. Biochem. Biophys.* **526,** 139–145; 2012), we calculate that antivenoms created from a mixture of recombinant antibodies could be produced on this scale for US$55–65 per gram.

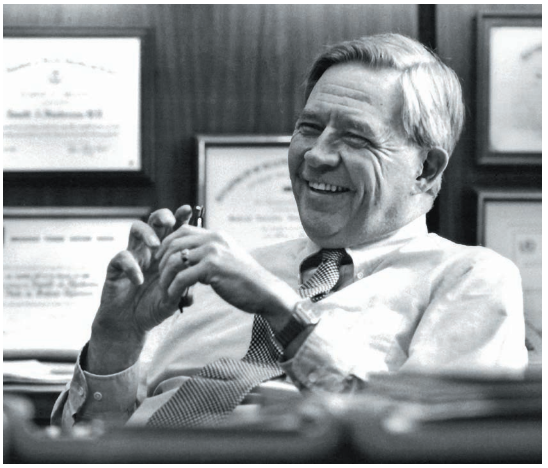
A typical African snakebite could therefore be treated with a pan-African recombinant-antibody antivenom for $30–150. This compares favourably with the wholesale cost of a typical dose of conventional antiserum ($60–600, which includes packaging and transport, as well as production, costs).

Andreas H. Laustsen* *Technical University of Denmark, Lyngby, Denmark.*
ahola@bio.dtu.dk
**On behalf of 4 correspondents (see go.nature.com/2dyztru for full list).*

Donald Ainslie Henderson

(1928–2016)

Epidemiologist who led the effort to eradicate smallpox.


COURTESY OF LEIGH HENDERSON

In the mid-1960s, smallpox was rampant in 31 countries in sub-Saharan Africa, Brazil and southeast Asia. Globally, between 10 million and 15 million cases were occurring each year, and one-third of those people were dying. Smallpox vaccines were of poor quality and attempts at disease surveillance wholly inadequate. In 1966, the World Health Assembly, the decision-making body of the World Health Organization (WHO), resolved to eradicate the disease. (A similar resolution had been passed in 1959, but it accomplished little, mainly because of a lack of funds and leadership.)

At the age of 38, Donald Ainslie Henderson became the head of the WHO's smallpox eradication programme. He was a revered leader for many of the 150,000 pox-warriors who marched until the last natural case of smallpox was diagnosed on 26 October 1977.

Henderson, who died on 19 August, was born in Lakewood, Ohio, in 1928 to an engineer and a nurse. He went to Oberlin College in Ohio and trained in medicine at the University of Rochester in New York, where he wrote a prizewinning paper about the 1832 epidemic of cholera in upstate New York.

In 1955, he joined the US Communicable Disease Center (now the Centers for Disease Control and Prevention, or CDC) where he was mentored by Alexander Langmuir. Langmuir had founded the Epidemic Intelligence Service (EIS) — in part to help the United States respond to biological threats. He was a demanding boss, advocating an on-the-ground approach, or "shoe leather" investigations and the importance of surveillance.

Henderson was appointed chief of the EIS programme, and later, chief of the CDC's surveillance section. In the lead-up to the 1966 WHO resolution, he'd been putting together a combined smallpox-eradication and measles-control programme involving 18 African countries. His leadership on this project resulted in him being assigned to the WHO's smallpox eradication programme.

Henderson set the vision of "smallpox zero". Faced with poor communication, civil wars, natural disasters and WHO bureaucracy — and before the existence of computers, mobile phones or fax machines — he had little choice but to delegate authority. And he was masterful at doing so.

The pox-warrior army included 812 staff from 73 countries. At the four-room nerve centre in Geneva, Switzerland, there were never more than ten staff members. The in-house rule was that all requests and letters had to be answered within two days. Realizing how important communication would be to the success of the effort, Henderson distributed more than 230 technical reports to keep the people involved in the programme abreast of developments. He even persuaded the WHO's *Weekly Epidemiological Record* to publish concise updates on smallpox incidence, problems and solutions. (Previously the bulletin had provided case totals with little interpretation or guidance.)

In 1967, Bill Foege, a CDC-trained missionary, showed in eastern Nigeria that vaccination of anyone who could have been exposed to the virus rather than whole populations could stop disease transmission. This 'ring' vaccination, the single-mindedness of fieldworkers and the concurrent invention of a two-pronged needle to simplify the vaccination procedure changed the course of the disease.

Outwardly, Henderson was very confident and optimistic about the eradication of smallpox — even when a helicopter and team members were captured by rebels in Ethiopia, civil war broke out in Pakistan, thousands of cases were discovered in Somalia, and the related human monkeypox surfaced in Zaire.

The public-health impacts of the smallpox-eradication programme are inestimable. In 1974, the WHO created an Expanded Programme on Immunization to roll out vaccination campaigns for other deadly infectious diseases. Henderson considered this to be the most important public-health legacy of the smallpox eradication programme.

In early 1977, when smallpox was endemic only in Somalia and the end was in sight, Henderson still had major careers ahead of him. He became dean of the School of Public Health at Johns Hopkins University in Baltimore, Maryland, a post he held for 13 years. He went on to serve in the US president's Office of Science and Technology and then in the Office of the Secretary of Health and Human Resources in Washington DC.

Then, in 1998, Henderson founded the Johns Hopkins Center for Civilian Biodefense Strategies (now the Center for Health Security at the University of Pittsburgh Medical Center in Pennsylvania). He was prescient. When anthrax spores were mailed to congressional and media offices after the 11 September attacks in 2001, Henderson was called on to give advice on bioterrorism preparedness. As concerns over global bioterrorism increased, he was asked to head the new Office of Public Health Emergency Preparedness.

D.A., as he was known since childhood, was tall and would command an audience with his stentorian voice. With ebullient hospitality and kindness, he and his wife Nana (an Oberlin classmate) welcomed scores of pox-fighters and friends to their home in Geneva, often with a steak on the grill and glass of single-malt whisky in hand.

In the policy world, he was not without controversy. He was sceptical of overly ambitious disease-eradication programmes. And to the end, he advocated strongly for getting rid of all remaining samples of the variola virus that causes smallpox. That view accords with the recommendations of a 1986 WHO advisory committee, and was supported by several national microbiological societies and many countries, but has not been supported by the US and Russian governments in periodic votes at the World Health Assembly.

When asked what should be eradicated next, D.A. would often respond, "bad management!". ■

Joel Breman *is senior scientist emeritus at the US National Institutes of Health in Bethesda, Maryland. He first met Henderson in 1967, before leaving for Guinea to work on the CDC-supported smallpox–measles programme. He later worked with him on certifying global eradication and on poxvirus research.*
e-mail: bremanj@mail.nih.gov

For News & Views online, go to
nature.com/newsandviews

CATALYSIS

Cobalt gets in shape

Solid cobalt-based catalysts are used commercially to convert carbon monoxide and hydrogen into synthetic fuels. It emerges that much more valuable chemicals can be produced by using a different form of cobalt catalyst. SEE LETTER P.84

MICHAEL CLAEYS

Catalysis lies at the heart of the production of more than 80% of all chemicals and petrochemicals[1]. In many processes, the chemical transformation takes place on the surface of catalytic, nanometre-scale metal particles. The size and shape of these particles and their chemical composition can greatly influence the effectiveness of a reaction and determine which products form[2,3]. Zhong *et al.*[4] report on page 84 that nanoprisms of a carbon-bearing cobalt compound, cobalt carbide (Co_2C), convert a mixture of carbon monoxide and hydrogen to valuable chemicals known as short-chain olefins. The authors' discovery is particularly surprising because the spherical counterparts of these crystallites are of little use in this reaction, and because spherical particles of cobalt metal produce a completely different product.

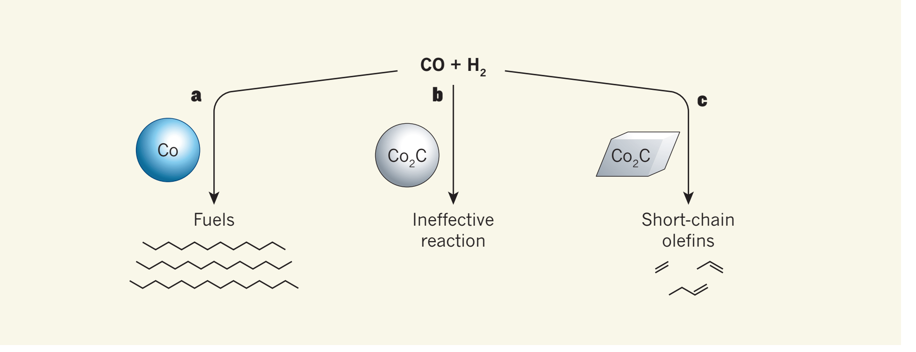


Figure 1 | Reaction selectivity in the Fischer–Tropsch process. a, In this process, synthesis gas — a mixture of carbon monoxide and hydrogen — can be converted to fuels in the presence of spherical nanoparticles of metallic cobalt. **b**, Spherical particles of cobalt carbide (Co_2C) do not effectively catalyse the formation of useful products from synthesis gas. **c**, Zhong *et al.*[4] report that nanoprisms of cobalt carbide selectively convert synthesis gas into compounds called short-chain olefins, which are high-value intermediates used to make polymers and other petrochemicals.

Mixtures of carbon monoxide and hydrogen are known as synthesis gas, and can be produced from a variety of carbon-containing feedstocks, including coal, natural gas, biomass and even waste. Synthesis gas is an important intermediate that can be further processed to yield chemicals such as methanol, ammonia and hydrogen, which are widely used in the chemical industry. Notably, it can also be used in the Fischer–Tropsch process, which converts the gas mixture into gaseous, liquid and solid hydrocarbons in a polymerization reaction that takes place on the surface of metallic cobalt or iron-based catalysts. The products contain linear chains of carbon atoms — the longest can contain more than 100 — and are used mainly to make high-quality synthetic fuels[5] (Fig. 1).

The Fischer–Tropsch process is practised industrially by several major companies around the world, but its commercial viability depends on many factors, most notably the price of crude oil. A potential way to greatly increase the profitability of this process is to capitalize on its ability to produce high-value chemicals — a seemingly underused feature. In particular, short-chain olefins (unsaturated hydrocarbons that contain two to four carbon atoms) can be made[5–7]. These compounds are widely used as building blocks for polymers, but are also used in solvents, cosmetics and detergents. They are normally derived from oil-based feedstocks, from minor components in natural gas or from methanol, and are produced in volumes that rank among the largest of any chemical product worldwide[6].

The maximum selectivity with which short-chain olefins can currently be obtained from an industrial Fischer–Tropsch operation is 24%, in a process[5] that uses an iron carbide catalyst. This process operates at a relatively high temperature of 330–350 °C, and is optimized for petrol production rather than olefin content. New catalysts need to be designed to maximize the selectivity with which these valuable olefins can be produced.

Highly promising results have previously been obtained[6–8] using iron-based catalysts modified with 'promoter' compounds, such as sulfur mixed with either potassium or sodium. Olefin selectivities of about 60% could be achieved in this way on a laboratory scale. However, these reactions normally need temperatures of 300–350 °C, which can cause carbon to be deposited on the catalyst surface, deactivating the catalyst and shortening its lifetime[6]. Short-chain olefins can also be selectively made from synthesis gas using an alternative[9] to the Fischer–Tropsch process that involves a different catalyst and minimizes undesired methane formation, but this method is not yet in commercial use.

Zhong *et al.* now report that nanoprisms of cobalt carbide produce short-chain olefins with up to about 61% selectivity in a Fischer–Tropsch reaction (Fig. 1), but at a much lower reaction temperature (250 °C) than is needed for iron-based catalysts. This finding is astonishing for several reasons. First, the industrial cobalt-catalysed Fischer–Tropsch reaction uses nanoparticles of metallic cobalt to produce synthetic fuels while minimizing olefin production (less than 5%)[10]. In that reaction, carbidic cobalt is typically viewed as an undesirable compound that has low catalytic activity, and which produces a large amount of unwanted methane[11].

Second, the industrial cobalt-catalysed reaction involves mostly spherical nanoparticles of cobalt. The observation that selective olefin formation occurs on differently shaped cobalt carbide crystallites is therefore unexpected, and underlines the role that different types of crystal surface have in catalytic reactions. Such a role is supported by Zhong and colleagues' theoretical predictions and by those of others[2].

But the most unexpected aspect of the authors' work is the discovery of the cobalt carbide nanoprisms themselves. These developed from mostly spherical, partially chemically reduced precursors made of a cobalt-manganese oxide composite during exposure to the reaction conditions. Large amounts of carbides do not normally form under these

conditions, and the authors show that the presence of manganese and residual sodium in the catalyst precursor might have been instrumental in the carbide formation and in causing the nanoprism shape to develop.

Some of these findings seem to be serendipitous, but their potential impact cannot be overestimated: they might open up pathways for the development of greatly improved systems for producing valuable chemicals from a variety of carbon sources. The findings also stimulate questions and ideas about the general role of cobalt carbides in the Fischer–Tropsch reaction, and about whether other shapes of cobalt and cobalt carbide nanoparticles are suitable for this process. Other forms of cobalt that have not conventionally been used in this reaction, such as cobalt nitrides, could also now be investigated.

Zhong and colleagues' cobalt-based catalyst might outperform its iron-based counterparts, because it operates at lower temperatures and is therefore potentially deactivated more slowly. However, the preparation of the catalyst is yet to be optimized, as is its formulation — the addition of promoter compounds might increase its performance, for example. Only about 30% of the carbon monoxide used in the reaction is currently converted, and so the operating conditions should also be studied to improve this, while maintaining or increasing the olefin selectivity.

Nonetheless, this is a groundbreaking contribution that further unlocks the immense potential of the Fischer–Tropsch process for producing chemicals. Zhong *et al.* have thrown open the reaction's treasure chest, and added fresh momentum to research into methods for making olefins from synthesis gas. Perhaps other valuable compounds, such as those that contain oxygen or nitrogen[12,13], are also within our grasp. ■

Michael Claeys *is at the Catalysis Institute and the DST/NRF Centre of Excellence in Catalysis (c*change), University of Cape Town, Rondebosch 7701, South Africa.*
e-mail: michael.claeys@uct.ac.za

1. Noyori, R. *Nature Chem.* **1,** 5–6 (2009).
2. Van Santen, R. A. *Acc. Chem. Res.* **42,** 57–66 (2009).
3. Somorjai, G. A. & Carrazza, J. *Ind. Eng. Chem. Fundam.* **25,** 63–69 (1986).
4. Zhong, L. *et al. Nature* **538,** 84–87 (2016).
5. Van de Loosdrecht, J. *et al.* in *Comprehensive Inorganic Chemistry II* Vol. 7 (eds Reedijk, J. & Poeppelmeier, K.) 525–557 (Elsevier, 2013).
6. Torres Galvis, H. M. & de Jong, K. P. *ACS Catal.* **3,** 2130–2149 (2013).
7. Torres Galvis, H. M. *et al. Science* **335,** 835–838 (2012).
8. Botes, G. F. *et al. Catal. Today* **275,** 40–48 (2016).
9. Jiao, F. *et al. Science* **351,** 1065–1068 (2016).
10. Dry, M. E. *Stud. Surf. Sci. Catal.* **152,** 196–257 (2004).
11. Claeys, M. *et al. J. Catal.* **318,** 193–202 (2014).
12. Pei, Y.-P. *ACS Catal.* **5,** 3620–3624 (2015).
13. Sango, T. *et al. Appl. Catal. A* **502,** 150–156 (2015).

CANCER

Acidic shield puts a chink in p53's armour

Underactivity of the transcription factor p53 can lead to tumour development. The discovery that the SET protein binds to and inhibits p53 points to a way to unleash the tumour suppressor's activity. SEE LETTER P.118

MICHELLE C. BARTON

In human history, possession of unbridled power has typically ended badly. The same is true in biology, as illustrated by studies of the tumour-suppressor protein p53. Stringent control is required to restrain this transcription factor's potent ability to cause cell death, arrest a cell in stasis or alter the course of metabolism. However, these controls need to be reversible, because p53 must be rapidly activated to protect cells from a wide variety of cellular stresses that promote tumour development[1]. On page 118, Wang *et al.*[2] reveal a previously unknown mechanism of restraining p53, which involves the formation of a reversible, acidic protein 'shield' that prevents the carboxy-terminal end of p53 from interacting with the cell's transcriptional machinery.

The carboxy-terminal domain (CTD) of p53 is a veritable hub of regulatory signalling. The six lysine residues within this 30-amino-acid region can be modified by several different types of molecule to alter how p53 regulates target genes, the stability of the protein, or its interactions with target DNA[1]. For instance, the addition of acetyl molecules to these lysine residues in response to cellular stressors such as DNA damage activates p53, leading to the transcription of target genes. But despite much research, exactly how this lysine acetylation controls p53's activity has remained unclear.

To drill down into this question, Wang and colleagues began with an unbiased, biochemical approach to identify proteins that interact with the p53 CTD, both when the protein is activated by lysine acetylation and when it lacks acetyl groups and is inactive. Surprisingly, and in contrast to previous studies, the authors found no proteins that bound to the acetylated CTD under their assay conditions, and only one, the tumour-promoting SET protein, that interacted with the unacetylated CTD.

The researchers show that SET acts as a transcriptional co-repressor, inhibiting p53's transcription-factor activity when bound to the CTD (Fig. 1). This inhibition relies on reversible electrostatic interactions between positively charged, basic amino acids in the unacetylated p53 CTD and a region of SET comprising long stretches of clustered, highly acidic and negatively charged amino acids.

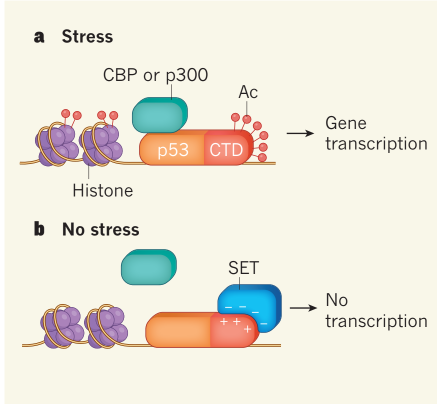


Figure 1 | A shield model of p53 regulation. **a**, Under conditions of cellular stress, acetyl groups (Ac) are added to six lysine amino-acid residues in the carboxy-terminal domain (CTD) of the tumour-suppressor protein p53. The protein binds to target DNA sequences and interacts with one of the two co-activator proteins CBP and p300, which acetylate DNA-associated histone proteins. These interactions together promote gene transcription. **b**, Wang *et al.*[2] report that, in the absence of stress and lysine acetylation, a highly acidic, negatively charged domain of the protein SET binds to the positively charged p53 CTD. Although SET-bound p53 can bind DNA, it cannot interact with p300 or CBP, and thus transcription is inhibited.

SET–CTD binding does not disrupt p53's interaction with its target DNA-binding sites, meaning that inactive p53 is poised to activate target genes, which is probably beneficial for a rapid stress response. Instead, SET acts as a shield, preventing the transcriptional co-activator proteins p300 and CBP from interacting with p53 and with nearby DNA and associated histone proteins, and so blocking target-gene activation in the absence of cellular stress.

Wang *et al.* defined highly acidic domains, such as that described for SET, as stretches of at least 46 amino acids, of which more than 76% of residues are acidic and are found in clusters across the domain. The authors searched the UniProt database[3] for other highly acidic domain proteins, and found only 49 that fitted these criteria, including the p53-interacting proteins DAXX, PELP1 and VPRBP. The group demonstrated that these proteins can bind to the unacetylated, but not the acetylated, p53 CTD. This finding suggests a

broad regulatory role for highly acidic domain proteins in an acetylation switch network, which probably extends beyond p53. However, it is puzzling that these proteins were not identified in Wang and colleagues' original screen. Moreover, it is difficult to reconcile the researchers' shield model of acidic-protein-mediated p53 inhibition with previous characterizations of DAXX (ref. 4) and PELP1 (ref. 5) as stress-dependent co-activators of p53.

The physiological importance of interactions between regulatory proteins and the p53 CTD has been established by engineering mice lacking this domain, which die within two weeks of birth[6,7]. Wang *et al.* mutated the six lysine residues in the p53 CTD to glutamines, which mimic the charge and structure of acetylated lysine and so effectively model permanent lysine acetylation. As such, mice harbouring this mutation lack SET binding to the CTD. These animals died within one day of birth, owing to unchecked cell death in the brain and severe neurological defects, underscoring the need for tight control of p53 activity during embryonic development.

By contrast, it has been shown[8] that replacement of lysine with arginine, which mimics a total lack of lysine acetylation — and, presumably, constitutive SET binding — produces no developmental anomalies. To confirm that these effects are attributable to SET, rather than to other highly acidic domain proteins, the authors deleted the mouse gene that encodes SET, which caused embryonic defects and death just before or after birth. Further studies are needed to determine whether this lethality results solely from unchecked p53 activation, or whether other functions of SET are also involved.

SET is a known tumour-promoting protein, and is aberrantly expressed in various cancers of the blood[9] and in solid tumours[10]. Previous studies of SET (for example, ref. 10) have focused mainly on its role as an inhibitor of protein phosphatase 2A (PP2A) — a tumour-suppressor protein that represses multiple signalling pathways that are aberrantly activated in many cancers, including the c-Myc, Wnt and PI3K/Akt pathways. Thus, therapies that inhibit SET may offer opportunities to treat cancer beyond simply unleashing p53. But such treatments must also take into consideration the complex consequences of altering SET activity.

In support of the therapeutic potential of targeting SET, Wang *et al.* showed that inhibition of SET production in mice led to regression of tumours with normal p53 levels, but not of tumours lacking the protein. However, concerns remain. For instance, tumours frequently harbour single-nucleotide mutations that alter the amino-acid sequence of p53 and so lead to production of a mutant protein. Disrupting SET–p53 interactions in cells carrying such mutations might lead to activation of a mutant protein that has deleterious tumour-promoting activities.

Profiling of the genomic regions with which SET is associated is now needed to determine: the breadth of p53-regulated genes affected by SET; whether SET's role is restricted to specific developmental stages or tissues; and whether p53 mutations that are implicated in cancers alter SET control and response. Moreover, studies that used SET inhibitors to increase PP2A activity in cancer[10] should be reinterpreted in light of the newly revealed role of SET as a protein shield. Combining SET inhibitors with drugs that inhibit lysine deacetylation[11] may offer effective therapeutic strategies in cancer treatment. ■

Michelle C. Barton *is in the Department of Epigenetics and Molecular Carcinogenesis, University of Texas MD Anderson Cancer Center, Houston, Texas 77030, USA.*
e-mail: mbarton@mdanderson.org

1. Wasylishen, A. R. & Lozano, G. *Cold Spring Harb. Perspect. Med.* **6,** a026211 (2016).
2. Wang, D. *et al. Nature* **538,** 118–122 (2016).
3. The UniProt Consortium. *Nucleic Acids Res.* **43,** D204–D212 (2015).
4. Lin, S.-C. & Li, Q. *Cell Res.* **17,** 301–302 (2007).
5. Nair, B. C. *et al. Cell Death Differ.* **21,** 1409–1418 (2014).
6. Hamard, P. J. *et al. Genes Dev.* **27,** 1868–1885 (2013).
7. Simeonova, I. *et al. Cell Rep.* **3,** 2046–2058 (2013).
8. Krummel, K. A., Lee, C. J., Toledo, F. & Wahl, G. M. *Proc. Natl Acad. Sci. USA* **102,** 10188–10193 (2005).
9. Christensen, D. J. *et al. Blood* **118,** 4150–4158 (2011).
10. Janghorban, M. *et al. Proc. Natl Acad. Sci. USA* **111,** 9157–9162 (2014).
11. Olzscha, H., Sheikh, S. & La Thangue, N. B. *Crit. Rev. Oncogen.* **20,** 1–17 (2015).

This article was published online on 14 September 2016.

BIOGEOCHEMISTRY

Rebalancing the global methane budget

A database of the carbon-isotope 'fingerprints' of methane has been used to constrain the contributions of different sources to the global methane budget. The surprising results have implications for climate prediction. SEE LETTER P.88

GRANT ALLEN

Globally averaged concentrations of atmospheric methane, a potent greenhouse gas, continue to rise. Explaining this trend by accurately accounting for sources and sinks of atmospheric methane gas remains a key challenge in climate science. On page 88, Schwietzke *et al.*[1] account for methane sources on the basis of a new carbon-isotope database. Their findings suggest that methane emissions associated with fossil-fuel use and production might be 20–60% higher than in current global inventories.

Methane is the second-largest contributor to climate radiative forcing — the change in energy trapped in the atmosphere as a result of greenhouse-gas emissions — and has a global-warming potential 28–34 times that of carbon dioxide (by equivalent mass) over a 100-year time frame[2]. So although average atmospheric methane concentrations are about 200-fold smaller than those of CO_2, understanding the causes of increases in global methane concentration is just as important as understanding those for increasing CO_2 levels, to aid climate prediction and inform emissions-reduction policy.

Atmospheric methane concentrations have been rising since the Industrial Revolution. A hiatus in this rise occurred between 1999 and 2006, although there is little consensus on the possible reasons for this — which vary from a reduction in coal mining and gas-industry emissions, especially in the countries of the former Soviet Union (see refs 3–5, for example), to the offsetting of increased anthropogenic emissions by decreasing wetland emissions[4,6]. Other studies have attributed the hiatus, at least in part, to changes in chemical species (reactive sinks) in the environment that react with methane[7,8], to reduced emissions from rice paddies[9] or simply to a plateauing of emissions from microbial and fossil-fuel sources[10]. The range of competing explanations exemplifies the complexity and uncertainty of balancing the global methane budget. But one thing is clear: methane levels have since vigorously resumed their upward trend[11], attracting strong and renewed scientific interest.

The scale of efforts to understand the most recent increases reflects the fact that researchers cannot easily explain the observed trend by comparing the rate of emissions with the rate at which methane is expected to be lost from the atmosphere through chemical reactions that occur in the environment. But the race to close the current global methane budget is just a sprint. A marathon effort is also required to take into account predicted changes in future human activity and the potential consequences

of global warming — such as the extensive release of methane from huge reservoirs currently trapped in permafrosts and ocean sediments (where it is stored as methane hydrate), and the bioclimatic response of wetlands, which are a major natural source of methane emissions.

One effective way to quantify the individual contributions made by the huge number of sources is to examine the isotopic fingerprint of methane molecules imparted by their source. The relatively short lifetime of methane in the atmosphere (about 12 years[2]) means that measured global patterns of the gas's carbon-isotope composition faithfully represent the average of recent inputs from the various sources, therefore allowing emission rates to be quantified by source type[4,12]. For example, thermogenic sources (those associated with fossil-fuel production and use) are enhanced in carbon-13 relative to biogenic (microbial) sources (Fig. 1).

Schwietzke and colleagues have compiled a database of previously measured carbon-13 methane isotopologues (methane molecules that contain carbon-13 instead of the more usual carbon-12) for principal source types, both biogenic and thermogenic. This is the largest database of its kind. It includes the statistical uncertainty in isotopic composition for each source type, based on available measurements, which allows constraints to be placed on estimates of emissions budgets using 'box' models.

The authors combined their database with previously reported global methane and methane-isotopologue measurements taken over the past three decades, and used the data in box modelling to show that total methane emissions associated with fossil-fuel use and production (including seepage from geological sources of fossil fuels) have not increased significantly over this time. However, such thermogenic emissions might currently be 60–110% higher than previous estimates, and might have been so for the past 30 years. After accounting for geological seepage, emissions attributable directly to the global fossil-fuel industry (natural gas, oil and coal production) are 20–60% higher than in current global inventories. One, possibly positive, implication of this analysis is that methane emissions associated with natural-gas production might have declined from about 8% of the total produced volume to about 2% between 1985 and 2013.

If Schwietzke and co-workers' findings are reinforced by similar studies and others using alternative methods, then there are several implications. First, emissions scenarios currently used for climate prediction need to be reassessed taking into account revised values for anthropogenic methane emissions. Second, the infrastructure for natural-gas production has become less 'leaky' over time, which has implications for emissions-weighted policies aimed at mitigating climate change. And third,

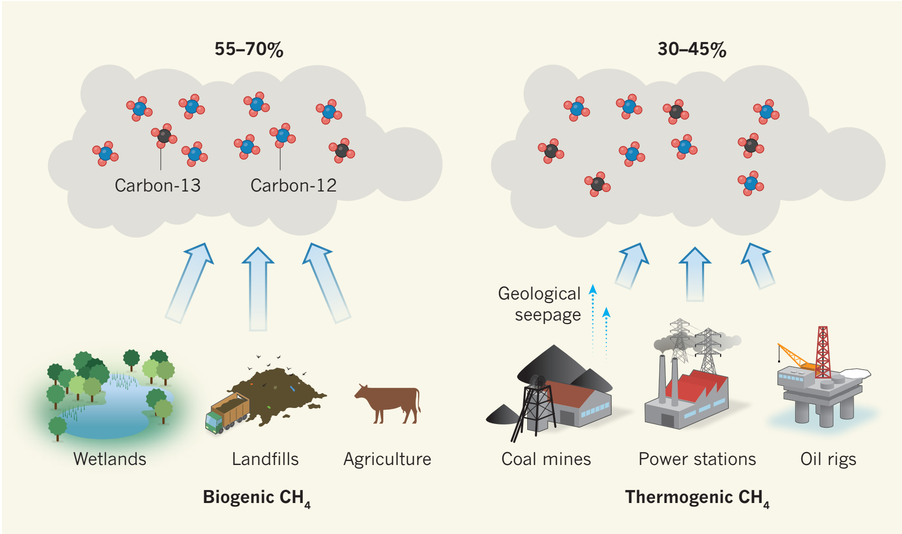


Figure 1 | Contributors to atmospheric methane. Methane produced from biogenic sources, such as wetlands, landfill sites and agriculture, contains less of the isotope carbon-13 than does methane from thermogenic sources (those associated with fossil-fuel extraction and use). Naturally occurring seepage from rocks is another thermogenic source, and is often associated with fossil-fuel extraction. Schwietzke *et al.*[1] have compiled a database of the carbon-isotope 'fingerprints' of different methane sources, and have used it to constrain the contributions of biogenic and thermogenic sources to the global atmospheric methane budget. The percentages shown were calculated (by G.A.) from data presented by Schwietzke and colleagues and from other data[4,7,14–17] used in their study, and are rounded to the nearest 1%. The ratio of carbon-12 to carbon-13 depicted in the clouds is illustrative, and does not precisely reflect experimental data.

more research on geological seepage might be needed.

However, Schwietzke and colleagues' conclusions are not without question or conflict. They markedly disagree with a range of 'flux-inversion' studies[4,6,13–16], which spatially attribute and optimally estimate a posteriori methane flux using reverse-transport modelling and a priori emissions inventories. Collectively, such studies have estimated a much lower emission rate (about 90 teragrams per year; 1 teragram is 10^{12} grams) than that reported by Schwietzke *et al.* for industrial fossil-fuel sources (approximately 155 Tg yr^{-1}). Moreover, Schwietzke and colleagues estimate that microbial methane emissions for 1985–2013 were about 15–33% lower than was reported in previous studies, but have been rising as a proportion of the total since 2001. The suggestion that microbial emissions have been increasing since the turn of the century is supported by another recent high-impact study[17] that also uses box modelling and isotopic fingerprinting. In other words, Schwietzke *et al.* rebalance the current global methane budget towards fossil fuels at the expense of biogenic emissions, although biogenic sources remain the dominant (and increasing) source.

The authors argue, and I agree, that known problems with a priori constraints and undersampling in key source areas such as tropical wetlands might lead flux-inversion models to amplify biogenic sources artificially, especially in the tropics. This is because inversion algorithms typically dump any uncertainty where they are least constrained by prior knowledge. Other systematic errors associated with the choice of data set and the way in which methane transport through the atmosphere is modelled may also be convolved in such inversions. Conversely, Schwietzke and co-workers' isotopic database, although useful and extensive, is only as good as its representation of sources, which depends on available sampling. But the abundance of carbon-13 in methane from different fossil reservoirs varies widely[18], and can even change within an individual reservoir as fossil fuels are extracted, especially in shale reservoirs.

So until there is convergence (within error) between inversion and box-modelling studies, the jury might still be out about the balance of the global methane budget. Such convergence will probably come from both directions: more-extensive sampling at sources would help to update the isotopologue database, improving box modelling; and better ambient sampling in key source regions such as the tropics would better constrain fluxes derived from flux-inversion models. Moreover, case studies of methane emissions on all spatial and temporal scales are integral to parameterizing emission processes for future climate predictions (see refs 19 and 20, for example). Fortunately, further sampling of methane and its isotopologues is on the horizon through simultaneous measurement programmes soon to get under way in the United States and Europe. ■

Grant Allen *is at the School of Earth and Environmental Sciences, University of*

Manchester, Manchester M13 9PL, UK.
e-mail: grant.allen@manchester.ac.uk

1. Schwietzke, S. *et al. Nature* **538,** 88–91 (2016).
2. Myhre, G. *et al.* in *Climate Change 2013: The Physical Science Basis. Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Stocker, T. F. *et al.*) Ch. 8 (Cambridge Univ. Press, 2013).
3. Dlugokencky, E. J. *et al. Geophys. Res. Lett.* **30,** 1992 (2003).
4. Chen, Y.-H. & Prinn, R. G. *J. Geophys. Res. Atmos.* **111,** D10307 (2006).
5. Simpson, I. J. *et al. Nature* **488,** 490–494 (2012).
6. Bousquet, P. *et al. Nature* **443,** 439–443 (2006).
7. Rigby, M. *et al. Geophys. Res. Lett.* **35,** L22805 (2008).
8. Montzka, S. A. *et al. Science* **331,** 67–69 (2011).
9. Kai, F. M., Tyler, S. C., Randerson, J. T. & Blake, D. R. *Nature* **476,** 194–197 (2011).
10. Levin, I. *et al. Nature* **486,** E3–E4 (2012).
11. Nisbet, E. G., Dlugokencky, E. J. & Bousquet, P. *Science* **343,** 493–495 (2014).
12. Quay, P. *et al. Glob. Biogeochem. Cycles* **13,** 445–461 (1999).
13. Fletcher, S. E. M., Tans, P. P., Bruhwiler, L. M., Miller, J. B. & Heiman, M. *Glob. Biogeochem. Cycles* **18,** GB4004 (2004).
14. Fraser, A. *et al. Atmos. Chem. Phys.* **13,** 5697–5713 (2013).
15. Wang, J. S. *et al. Glob. Biogeochem. Cycles* **18,** GB3011 (2004).
16. Bergamaschi, P. *et al. J. Geophys. Res.* **114,** D22301 (2009).
17. Schaefer, H. *et al. Science* **352,** 80–84 (2016).
18. Stolper, D. A. *et al. Geochim. Cosmochim. Acta* **161,** 219–247 (2015).
19. Myhre, C. L. *et al. Geophys. Res. Lett.* **43,** 4624–4631 (2016).
20. O'Shea, S. *et al. Atmos. Chem. Phys.* **14,** 13159–13174 (2014).

PARTICLE PHYSICS

Search for neutrinoless double-β decay

Neutrinos are much lighter than the other constituents of matter. One explanation for this could be that neutrinos are their own antiparticles and belong to a new class of 'Majorana' particle. An experiment sets strong constraints on this scenario.

GIORGIO GRATTA

The surprising discovery that elementary particles called neutrinos oscillate[1] earned its finders the 2015 Nobel Prize in Physics. Neutrinos come in three 'flavours' and, as they travel, their flavour can change. These oscillations are a purely quantum-mechanical phenomenon and can occur only if neutrinos have mass. However, we know from various observations[2] that these masses must be minuscule, probably less than 1 electronvolt. By comparison, the next-lightest particle, the electron, weighs about half a million electronvolts.

The smallness of neutrino masses might be explained if neutrinos are Majorana particles — that is, indistinguishable from their antiparticles. This is possible because neutrinos are electrically neutral. All other fundamental particles of matter, such as electrons and quarks, have an electric charge that clearly distinguishes them from their antiparticles. The Majorana explanation could be confirmed through the observation of a radioactive decay process called neutrinoless double-β decay. However, writing in *Physical Review Letters*, the KamLAND-Zen Collaboration[3] finds no evidence for this process, suggesting that, if it exists, it is even rarer than previously known.

Conventional (two-neutrino) double-β decay is not particularly remarkable — it is a process whereby the nuclei of certain isotopes decay and emit two electrons and two neutrinos (Fig. 1a). However, neutrinoless double-β decay, in which no neutrinos are emitted, could occur only for Majorana neutrinos (Fig. 1b). In this case, the particle–antiparticle nature would be blurred and a neutrino could be emitted and reabsorbed in the same elementary process.

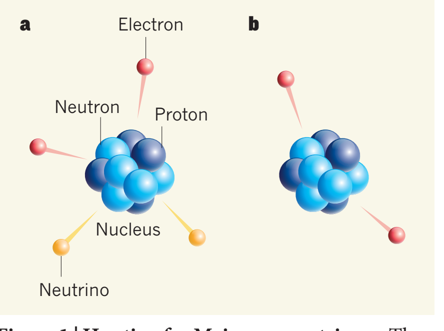


Figure 1 | Hunting for Majorana neutrinos. The KamLAND-Zen Collaboration[3] sets the strongest limits so far on the rate of a radioactive-decay process called neutrinoless double-β decay. **a,** In conventional double-β decay, a nucleus emits a pair of electrons and a pair of neutrinos. **b,** If neutrinos are Majorana particles (indistinguishable from their antiparticles), neutrinoless double-β decay can also occur, in which only two electrons are emitted from the nucleus.

Neutrinoless double-β decay is possible only if neutrinos have mass, a condition that has now been confirmed, thanks to the detection of neutrino oscillations. The race is therefore on to find evidence for this elusive process. Like everything else that involves neutrinos, this is not easy. The smallness of neutrino masses guarantees that the decay, if it exists, is extremely rare — in other words, the half-life of the candidate nucleus is exceedingly long. Neutrinoless double-β-decay experiments therefore observe a large quantity of a candidate isotope (a few hundred kilograms in the present generation of experiments) in the hope of seeing a handful of decay processes in which only two electrons are emitted, each having a kinetic energy of about one megaelectronvolt (ref. 4).

However, the required isotopes are often rare in nature and need to be separated from other isotopes of the same element, which is already quite a complex enterprise. Furthermore, experimental 'backgrounds' can produce electrons in the megaelectronvolt energy range that look similar to those expected from neutrinoless double-β decay. In particular, there are major backgrounds from cosmic rays and the natural radioactivity of elements such as uranium and thorium — elements present in Earth's crust that cause unavoidable contamination of all matter around us.

The strategy adopted for suppressing these backgrounds has evolved over the past 30 years. For example, some initial experiments fashioned small quantities of the isotope (in the gram-to-kilogram range) into large, extremely thin sheets, and analysed the energy and momentum of the two emitted electrons separately. Although this method is superb at distinguishing signal from background, it is too expensive if larger quantities of the isotope are required. Therefore, most modern detectors use the isotope in bulk — in either solid or liquid form[4].

Another experimental consideration is that neutrinoless double-β decay would produce two electrons with a fixed (and known) combined energy, whereas background events mostly produce electrons with a wide range of energies. As a result, the energy resolution of the detectors in these experiments has been considered, until now, the most crucial factor in distinguishing signal from background. The best resolution comes from crystals of tellurium dioxide ($^{130}TeO_2$) and semiconductors such as germanium (^{76}Ge). However, because these crystals are limited to a few kilograms in mass, the corresponding detectors consist of segmented arrays of crystals, and the surfaces and construction materials between the crystals produce additional backgrounds[4].

The approach used by the KamLAND-Zen authors is strikingly different. They use an isotope of xenon (^{136}Xe), dissolved in about 10^6 kg of a liquid scintillation material in a detector; the liquid scintillator emits light when it absorbs an energetic particle.

The detector's energy resolution is nowhere near that of the crystal detectors. However, its huge volume strongly shields the region of the detector in which the measurement is made from the external radioactive backgrounds, and the liquid scintillator can be purified to an extreme level[5]. As long as the authors can prove that there is no background, they can use a lack of signal for neutrinoless double-β decay to constrain the half-life of the process.

Indeed, the authors find no evidence for neutrinoless double-β decay, showing that, if the decay exists, its half-life must be longer than 1.07×10^{26} years, which is more than 7×10^{15} times the age of the Universe. Although this is a negative result, it is important because the possibility of discovering neutrinoless double-β decay — and therefore the Majorana nature of neutrinos — is one of the few opportunities we know of for finding evidence of physics beyond the standard model of particle physics, and potentially solving the puzzle of why neutrinos are so much lighter than all other matter particles.

Therefore, although the KamLAND-Zen result is impressive, physicists are devising improvements to the existing detection techniques, with the aim of building a new generation of detectors to extend the hunt for Majorana neutrinos. Current experiments use both segmented detectors, such as those involving crystals[6], and large homogeneous detectors, such as the liquid-scintillation detector of KamLAND-Zen and liquid xenon time-projection chambers[7]. But the strong constraints on neutrinoless double-β decay that are set by large homogeneous detectors might suggest that the future belongs to them.

Eventually, when the external backgrounds of large homogeneous detectors have been eliminated, conventional double-β decay will become the dominant background. This process produces electrons that are uniformly distributed in the detector in the same way as a potential signal, so increasing the detector size will not help to reduce this background. However, because conventional double-β decay leads to the production of two electrons that have a wide range of combined energies, it can be distinguished from the signal if the energy resolution of the detector is significantly better than that of KamLAND-Zen. Liquid xenon time-projection chambers, which have a resolution between that of the crystal and liquid-scintillation detectors, might be the detector of choice for future experiments. Alternatively, perhaps scientists will find a way to substantially improve the resolution of liquid-scintillation detectors, or grow really large crystals. ■

Giorgio Gratta *is in the Department of Physics, Stanford University, Stanford, California 94305, USA.*
e-mail: gratta@stanford.edu

1. Barger, V., Marfatia, D. & Whisnant, K. L. *The Physics of Neutrinos* (Princeton Univ. Press, 2012).
2. Olive, K. A. *et al.* (Particle Data Group) *Chin. Phys. C* **38,** 090001 (2014).
3. Gando. A. *et al.* (KamLAND-Zen Collaboration) *Phys. Rev. Lett.* **117,** 082503 (2016).
4. Giuliani, A. & Poves, A. *Adv. High Energy Phys.* **2012,** 857016 (2012).
5. Alimonti, G. *et al. Nucl. Instrum. Meth. Phys. Res. A* **609,** 58–78 (2009).
6. Agostini, M. *First Results from GERDA Phase II*; www.mpi-hd.mpg.de/gerda/public/2016/t16_neutrino_gerda_ma.pdf (2016).
7. EXO-200 Collaboration. *Nature* **510,** 229–234 (2014).

This article was published online on 21 September 2016.

HUMAN MIGRATION

Climate and the peopling of the world

The human dispersal out of Africa that populated the world was probably paced by climate changes. This is the inference drawn from computer modelling of climate variability during the time of early human migration. **See Letter p.92**

PETER B. DEMENOCAL & CHRIS STRINGER

One of the most puzzling questions about the origins of modern humans has been why the dispersal of *Homo sapiens* out of Africa occurred so long after their first known appearance in east Africa approximately 150,000 to 200,000 years ago[1]. Fossil, archaeological and genetic evidence indicates that early migrations out of Africa into the Levant (eastern Mediterranean) and the Arabian peninsula occurred around 120,000 to 90,000 years ago[1], but the further dispersal of our kind halfway around the world did not begin until about 60,000 years ago[1]. This out-of-Africa migration was pulsed, with waves of dispersal eastward to south Asia, Indonesia and Australia by 50,000 years ago, migration westward to Europe by 45,000 years ago[1], migration into north Asia by 20,000 years ago and to the Americas by 15,000 years ago[2] (Fig. 1). On page 92, Timmermann and Friedrich[3] provide modelling insights into the potential role of climate in the human migration out of Africa.

The role of climate change in pacing these ancient human dispersals has been the subject of intense study and debate. All hypotheses share the basic principle that climate affects resource richness, which, in turn, sets the 'carrying capacity' — the human population that can be supported in a given region. This then guides human dispersal. Climate agents that might affect resource richness include large volcanic eruptions[4], glacial 'Heinrich events' associated with ice-sheet collapse[5], orbital monsoonal-rainfall changes (Earth's orbit undergoes slight changes in its rotational axis every 21,000 years, which affects seasonal solar radiation and thus monsoonal climate)[6–8] and sea-level fluctuations[9].

Many studies have used climate models to explore the effects of these palaeoclimatic agents on human migrations[5,7,10]. These simulations provide spatio-temporal models of ancient climates that can be compared with the available fossil, archaeological and genetic evidence. The challenge has been to construct a model that has sufficiently realistic palaeoclimate representations, while simultaneously modelling changes in human carrying capacity that match observed dispersal routes and timing[7]. It's a tough problem.

Timmermann and Friedrich tackle this with the most comprehensive climate, vegetation and human-dispersal modelling study performed so far. They use a fully coupled ocean–atmosphere–vegetation climate model that is forced by specified changes in orbital insolation (solar-radiation levels that depend on Earth's tilt and changes in the Earth–Sun distance), carbon dioxide levels, glacial ice and sea-level boundary conditions to compute transient changes in climate and vegetation over the past 125,000 years. The authors validated the model climate fields against available palaeoclimate and palaeoceanographic data to ensure that the results were reasonable.

There are, however, some deficiencies in the model, such as the weaker-than-observed African monsoonal-rainfall response to orbital insolation forcing that is evident in nearly all such models[11]. The authors modelled human dispersals using computer simulations of population density as a function of environmental parameters (while also accounting for parameter uncertainties) at a global geographic resolution of 1° latitude × 1° longitude.

What Timmermann and Friedrich found was both remarkable and instructive. Today, the Sahara and Arabian deserts form an effective barrier to faunal dispersals out of Africa. But in the past, changes in the orientation of Earth's axis of rotation at that time invigorated the monsoonal climate and established wetter conditions in the Arabian and Sinai

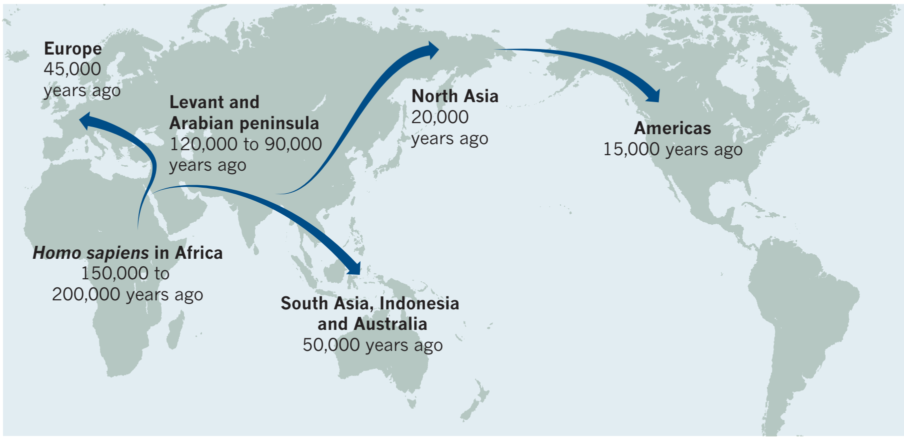


Figure 1 | Human migration out of Africa. Previous studies[1,2] of human migration out of Africa, using fossil, archaeological and genetic evidence, have provided a timeline of the human global dispersals shown. Timmermann and Friedrich[3] used linked climate, vegetation and human-dispersal models to understand how climate change may have paced the tempo of human migrations out of Africa. Their results support the view that climate may have been a key factor, but show both similarities and differences when compared with the results of previous studies. One notable difference is that Timmermann and Friedrich suggest a much earlier arrival of modern humans in Europe.

peninsulas, enabling migration paths out of Africa along vegetated, resource-rich corridors. These corridors were established during three time windows: 130,000 to 118,000 years ago, 106,000 to 94,000 years ago and 89,000 to 73,000 years ago (although the first green corridor, established 130,000 to 118,000 years ago, was not associated with human migration out of Africa in the authors' model). These age ranges coincide with warm substages within the single interglacial known as Marine Isotope Stage 5 (MIS5) of the geological temperature record. This orbital pacing of migration waves out of Africa supports earlier conclusions that the resulting environmental change was a probable mechanism that drew early populations of humans out of their ancestral African home because of the establishment of new, resource-rich exit routes[6–8,10,12,13].

However, the onset of dry, resource-poor conditions during glacial MIS4 (71,000 to 60,000 years ago) terminated the exchange like closing a valve. The next key migration wave out of Africa occurs during the subsequent, orbitally driven increase in monsoonal rainfall during early MIS3 (59,000 to 47,000 years ago). This wave of migration boosted remnant Eurasian populations, leading to rapid population increases in Europe and elsewhere between 60,000 and 40,000 years ago. At the same time, the authors simulate a rapid eastward expansion into India and south Asia, with humans arriving in Australia by 60,000 to 50,000 years ago. Migration into north Asia and then into the Americas occurs only when glacial conditions start to wane after around 20,000 years ago.

Timmermann and Friedrich explored the sensitivity of these model results to changes in several climate and dispersal parameters. They found that the orbital pacing of human dispersal events out of Africa is a robust result, as is the importance of MIS4 aridification in cutting off the exchange between the populations in northeastern Africa and the rapidly eastward-spreading group in southern Asia. The authors also show that millennial-scale climate oscillations, comparable to rapid warming or cooling episodes known as Dansgaard–Oeschger events, had little effect on migration times.

How well do the estimated migration-wave timings match previous archaeological, fossil and genetic data? For Arabia, archaeological evidence indicative of modern human presence does suggest modern human dispersals from Africa or the Levant between about 120,000 and 75,000 years ago)[9,12], whereas only the potentially oldest dating evidence for Skhul and Qafzeh (in the Levant)[13] falls earlier than the age ranges modelled by Timmermann and Friedrich (although an ancient jawbone found in Tabun Cave[14] raises the possibility that modern humans had an even earlier presence). For the Indian subcontinent, there is only limited archaeological evidence for a modern human presence before 50,000 years ago, although a partial cranium and jawbone are known from Laos at this time[15]. In China, there are several claims from fossil evidence for a modern human presence before 80,000 years ago, but these may require further confirmation[16].

The most obvious discrepancy in Timmerman and Friedrich's results is their suggestion that southern Europe experienced a low-density wave of occupation by modern humans before 80,000 years ago, which is more than 35,000 years earlier than the generally accepted evidence from archaeology and fossil remains[17]. The authors suggest that these earliest modern pioneers could have been assimilated by the more numerous Neanderthals. However, genetic signatures of these proposed early pioneers have not yet been detected in the genomes of subsequent Neanderthal individuals in Europe, and it could also be argued that, as with later Siberian and Romanian fossils[18], these early modern humans and their lineages simply went extinct. However, it seems unlikely that such early modern dispersals would not have left at least some distinctive archaeological traces, something that has not yet been detected.

Although such human–climate interactions may seem too complex to model with any fidelity, ancient population dynamics across north Africa provide an instructive example. Between 12,000 to 5,000 years ago, the vast Sahara was nearly completely vegetated with wooded grasslands, permanent lakes and rivers[19]. This region was alive with people and cultural activity until about 5,000 years ago, when the monsoon rains weakened and retreated as a result of changes in Earth's orbit. The archaeological record documents the massive and rapid depopulation of the north African interior around 5,000 years ago, at the same time as the establishment of the present-day Sahara Desert[20]. This well-documented case study illustrates just how effectively climate can shape life, including the peopling of the planet. ■

Peter B. deMenocal *is at the Lamont-Doherty Earth Observatory, Columbia University, Palisades, New York 10964, USA.*
Chris Stringer *is in the Department of Earth Sciences, Natural History Museum, London SW7 5BD, UK.*
e-mails: peter@ldeo.columbia.edu; c.stringer@nhm.ac.uk

1. Groucutt, H. S. *et al. Evol. Anthropol.* **24,** 149–164 (2015).
2. Goebel, T., Waters, M. R. & O'Rourke, D. H. *Science* **319,** 1497–1502 (2008)
3. Timmermann, A. & Friedrich, T. *Nature* **538,** 92–95 (2016).
4. Ambrose, S. H. *J. Hum. Evol.* **34,** 623–651 (1998).
5. Carto, S. L., Weaver, A. J., Hetherington, R., Lam, Y. & Wiebe, E. C. *J. Hum. Evol.* **56,** 139–151 (2009).
6. Castañeda, I. S. *et al. Proc. Natl Acad. Sci. USA* **106,** 20159–20163 (2009).
7. Eriksson, A. *et al. Proc. Natl Acad. Sci. USA* **109,** 16089–16094 (2012).
8. Osborne, A. H. *et al. Proc. Natl Acad. Sci. USA* **105,** 16444–16447 (2008).
9. Armitage, S. J. *et al. Science* **331,** 453–456 (2011).
10. Jennings, R. P. *et al. Quat. Int.* **382,** 181–199 (2015).
11. Pausata, F. S. R., Messori, G. & Zhang, Q. *Earth Planet. Sci. Lett.* **434,** 298–307 (2016).
12. Groucutt, H. S. & Petraglia, M. D. *Evol. Anthropol.* **21,** 113–125 (2012).
13. Grün, R. *et al. J. Hum. Evol.* **49,** 316–334 (2005).
14. Rak, Y., Ginzburg, A. & Geffen, E. *Am. J. Phys. Anthropol.* **119,** 199–204 (2002).
15. Demeter, F. *et al. PLoS ONE* **10,** e0121193 (2015).
16. Michel, V. *et al. J. Hum. Evol.* http://dx.doi.org/10.1016/j.jhevol.2016.07.008 (2016).
17. Svoboda, J. in *Emergence and Diversity of Modern Human Behavior in Palaeolithic Asia* (eds Kaifu, Y., Izuho, M., Goebel, T., Sato, H. & Ono, A.) 23–33 (Texas A&M Univ. Press, 2015).
18. Fu, Q. *et al. Nature* **534,** 200–205 (2016).
19. deMenocal, P. B. & Tierney, J. E. *Nature Educ.* **3,** 12 (2012).
20. Manning, K. & Timpson, A. *Quat. Sci. Rev.* **101,** 28–35 (2014).

This article was published online on 21 September 2016.

doi:10.1038/nature19773

Circuit-based interrogation of sleep control

Franz Weber[1] & Yang Dan[1]

Sleep is a fundamental biological process observed widely in the animal kingdom, but the neural circuits generating sleep remain poorly understood. Understanding the brain mechanisms controlling sleep requires the identification of key neurons in the control circuits and mapping of their synaptic connections. Technical innovations over the past decade have greatly facilitated dissection of the sleep circuits. This has set the stage for understanding how a variety of environmental and physiological factors influence sleep. The ability to initiate and terminate sleep on command will also help us to elucidate its functions within and beyond the brain.

Sleep is a seemingly unproductive behavioural state that takes up a large proportion of our lives, but insufficient sleep can profoundly impair our cognitive performance during wakefulness. Long-term sleep deprivation is also linked to many other health problems, including obesity and cardiovascular diseases. At the behavioural level, sleep has been observed widely across the animal kingdom, including in worms and flies, as well as vertebrates. However, the existence of two distinct types of sleep—rapid eye movement (REM) sleep and non-REM (NREM) sleep—was previously thought to be restricted to mammals and birds and has only recently been identified in reptiles[1].

Wakefulness, NREM sleep and REM sleep can be clearly distinguished based on electroencephalogram (EEG) and electromyogram (EMG) recordings, making sleep a directly quantifiable behaviour. During wakefulness, the EEG exhibits high-frequency, low-amplitude activity ('desynchronized EEG'), and the EMG shows high muscle tone (Fig. 1a, c). In contrast, the EEG during NREM sleep is dominated by high-amplitude, low-frequency (0.5–4.5 Hz) activity ('synchronized EEG') together with sleep spindles (waxing and waning of 9–15 Hz oscillations lasting for a few seconds). REM sleep is associated with vivid dreaming; it is also called paradoxical sleep, as it is characterized by desynchronized EEG resembling that during wakefulness, but the EMG shows a complete paralysis of postural muscles[2]. The proportions of time the animal spends in wakeful, NREM and REM states and the temporal patterns of state transitions vary widely across species[3] (Fig. 1b, d). However, there are some well-conserved features. For example, animals normally enter REM sleep from NREM sleep but not directly from wakefulness.

Up until the 20th century, sleep was believed to be a passive process, caused by reduced sensory stimulation that allows our normal mental and physical activities to shut down. We now know, however, that both NREM and REM sleep are controlled by distinct neural circuits in the brain, the malfunction of which causes a variety of sleep disorders. Since the discovery of the ascending reticular activating system more than half a century ago[4], we have learned a great deal about the neural circuits supporting wakefulness[5,6]. In contrast, the neural mechanisms generating sleep have been far more elusive. While studies based on lesion, electrical stimulation and pharmacological manipulations (Fig. 2a, b) have implicated multiple brain regions that are important for sleep[5,6], which neurons are responsible for triggering and maintaining NREM or REM sleep and how they are connected to each other remain largely unknown. A main difficulty resides in the fact that the sleep-promoting neurons are often spatially intermingled with, but outnumbered by, wake-promoting neurons, making it difficult to target them selectively for circuit analysis.

Over the past decade, several new techniques have become widely available, including optogenetics[7], pharmacogenetics[8], imaging with

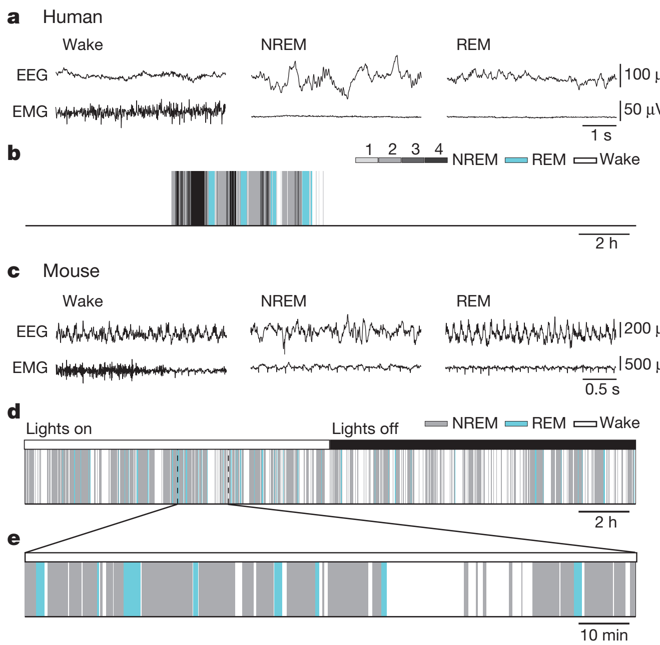


Figure 1 | Sleep in humans and mice. a, Examples of a human electroencephalogram (EEG) and electromyogram (EMG) recordings during wakefulness, NREM sleep (stage 3) and REM sleep. **b**, Colour-coded brain states (hypnogram) during a continuous 22-h recording from a healthy human subject. The EEG recordings (**a**) and hypnogram (**b**) are from the Sleep EDF database[141,142]. In humans, sleep is consolidated with rare awakenings during the night. REM sleep occurs regularly every ~90 min. **c**, Example EEG and EMG recordings from a mouse during wakefulness, NREM and REM sleep. **d**, Hypnogram during a continuous 24 h recording from a mouse in a dark–light cycle. Mice sleep more during the light cycle. Compared to humans, mice exhibit fragmented sleep patterns, characterized by short sleep bouts and frequent awakenings. **e**, A 2 h segment from the hypnogram in **d** shown at an expanded scale. In mice, REM sleep occurs every 10 to 20 min.

[1]Division of Neurobiology, Department of Molecular and Cell Biology, Helen Wills Neuroscience Institute, Howard Hughes Medical Institute, University of California, Berkeley, California 94720, USA.

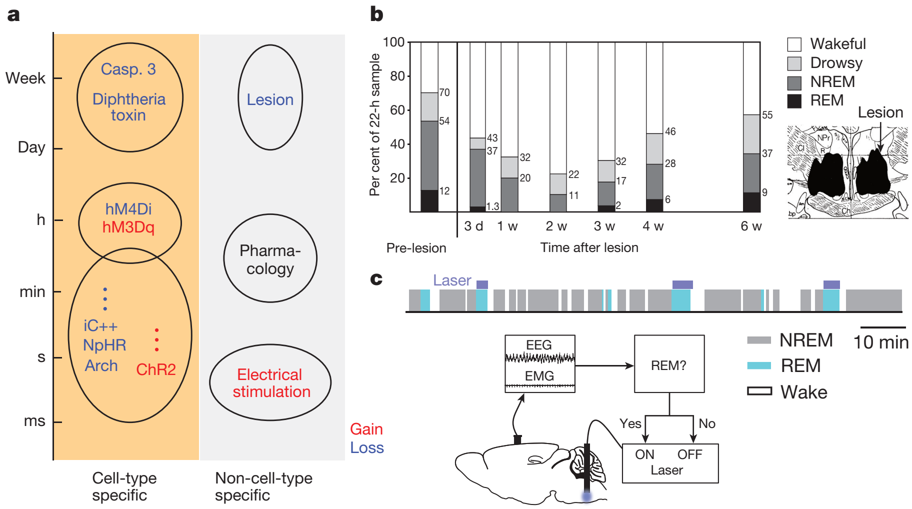


Figure 2 | Methods for neuronal manipulations. a, Various techniques grouped depending on whether they are cell-type-specific (orange box) or non-specific (grey box). Circle indicates timescale of each method. Text colour depicts activation versus suppression of neural activity (red and blue; gain and loss of function, respectively). Non-cell-type-specific methods include electrical stimulation of neurons or fibre tracts, pharmacological application of agonists or antagonists to specific receptors, and various methods for lesions. Using optogenetics, the light-activated cation-channel channelrhodopsin (ChR2) can be expressed in genetically defined cell types, allowing for their activation by light within milliseconds[7]. By contrast, light activation of the chloride pumps halorhodopsin (NpHR) or archaerhodopsin (Arch) causes rapid neural inhibition. Recently, a light-activated chloride channel (iC++) was developed[143]. Using pharamacogenetics, neurons can be continuously activated or inhibited for hours. The method relies on an extrinsic G protein-coupled muscarinic receptor, of which an excitatory (hM3Dq) and an inhibitory (hM4Di) version exist. The receptor is activated only by a physiologically inert, synthetic ligand[8]. Lesions of specific cell types can be achieved using genetically encoded toxins (diphtheria toxin)[144] or apoptotic signalling molecules (caspase 3)[145]. **b**, Lesion of the POA and BF in cats induced long-lasting inhibition of sleep (d, days; w, weeks). Right, schematic depicting the lesioned region. Data reprinted and adapted with permission from ref. 19. **c**, Millisecond precision of neural manipulation afforded by optogenetics allowed a closed-loop stimulation protocol to test the role of GABAergic ventral medulla neurons in REM sleep maintenance. The laser was turned on after spontaneous REM onset and turned off at the end of the REM episode. Data adapted from ref. 74.

genetically encoded calcium indicators[9] and virus-mediated circuit tracing[10,11] (Figs 2 and 3). Combined with mouse genetics, these techniques endow us with an unprecedented capability for measuring and controlling the activity of specific cell types and dissecting their synaptic connections, greatly facilitating our investigation of the mechanisms that control sleep. In this review, we focus on the neural circuits controlling both NREM and REM sleep in the mammalian brain, with a particular emphasis on studies enabled by recently developed technologies. The function and genetics of sleep and studies in non-mammalian species are not covered here, but can be found in several recent reviews[12–16].

Forebrain control of sleep versus wakefulness

Multiple brain areas have been implicated in controlling the switch between wakefulness and the general state of sleep, including both REM and NREM sleep. Many of these areas are located in the forebrain, including the preoptic hypothalamus, basal forebrain and lateral hypothalamus.

Preoptic hypothalamus

The preoptic area (POA) of the anterior hypothalamus has long been known to be important for sleep generation. In the 1920s, Von Economo found that damage to the POA was associated with insomnia in human patients[17]. By systematically varying the location of a brain lesion in the rat, Nauta concluded that the POA is a 'sleep center'[18], a notion that was supported by subsequent lesion[19,20] and muscimol injection[21] experiments in the cat. In the 1990s and 2000s, c-Fos immunohistochemistry following sustained sleep revealed sleep-active GABAergic neurons in the ventrolateral preoptic area (VLPO) and the median preoptic nucleus (MnPO)[22,23], and selective lesion of the VLPO drastically reduced NREM sleep[24]. A recent study showed that pharmacogenetic activation of the c-Fos-labelled neurons in the POA induces sleep, further supporting their causal role in sleep regulation[25].

Neurons in the VLPO and MnPO are likely to promote sleep through their inhibitory projections to wake-promoting brain areas (Fig. 4). Targets of these projections include the major wake-promoting monoaminergic centres such as the histaminergic tuberomammillary nucleus (TMN), serotonergic dorsal and median raphe nuclei (DRN and MRN) and noradrenergic locus coeruleus (LC)[26,27]. They also project to the perifornical lateral hypothalamus[28], which contains the wake-promoting orexin (also known as hypocretin) neurons[29], the ventral periaqueductal grey matter (vPAG)[30] and the parabrachial nucleus[26], which has also been shown to be important for wakefulness and arousal[31]. Among these targets, the POA projection to the TMN in the posterior hypothalamus appears particularly strong[26,27,32] and may powerfully inhibit TMN neurons during sleep[33]. The importance of this projection is supported by the observation that inactivating the posterior hypothalamus with muscimol injection can strongly promote sleep[21] and can reverse the insomnia induced by a POA lesion[20]. Microdialysis measurements showed that the extracellular concentrations of GABA in the LC and DRN also increase during sleep and that GABA levels are highest during REM sleep[34,35], when these monoaminergic neurons are virtually silent[36]. This shows the importance of GABAergic inputs in regulating the firing of these neurons across brain states, and it is probable that POA sleep-active neurons provide a substantial source of such GABAergic inputs. In addition to GABA, the neuropeptide galanin is also expressed in, and presumably released by, many POA sleep-active neurons[26,27,32], which can inhibit the TMN histaminergic neurons[37] and noradrenergic neurons of the LC[38].

What inputs activate the POA neurons during sleep and/or suppress them during wakefulness? Anatomically, histaminergic, noradrenergic and serotonergic axon fibres are observed in the POA[39]. *In vitro* recordings showed that ~70% of VLPO neurons are inhibited by noradrenaline and acetylcholine, with a subset also inhibited by serotonin[40]. Although histamine does not directly inhibit VLPO neurons, it might activate noradrenaline-excited interneurons, which could in turn inhibit the sleep neurons[41]. These effects of the wake-promoting neuromodulators probably suppress the activity of POA sleep neurons during wakefulness.

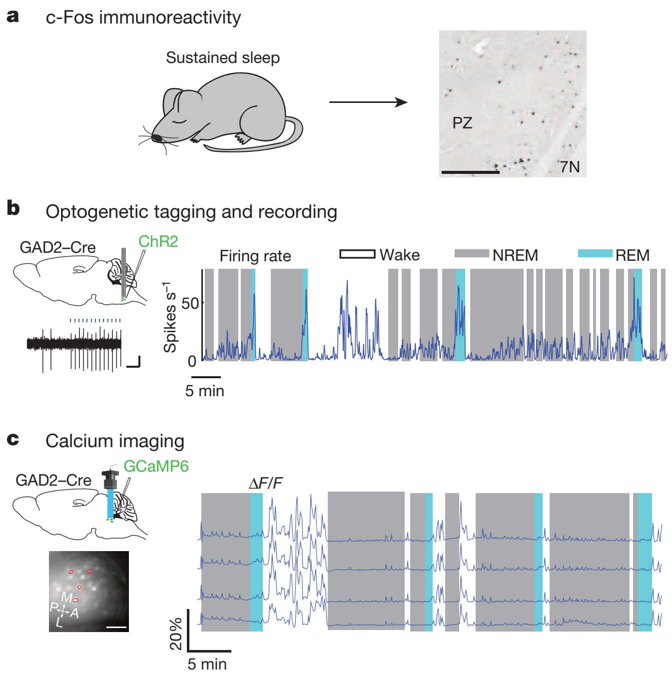


Figure 3 | Methods for measuring neural activity. **a**, c-Fos immunohistochemistry is widely used to detect sleep-active neurons. Following spontaneous sleep or deprivation-induced sleep rebound, brain tissue is stained for the expression of the immediate early gene *c-Fos*, used as a marker for neuronal activation. Right, brain section showing c-Fos-positive cells (black) in the parafacial zone (PZ; 7N, facial nucleus; scale bar, 300 μm). Data reprinted with permission from ref. 76. **b**, Recording from genetically defined cell types using optogenetic tagging. Left, recording with optrodes (microelectrodes coupled with an optic fibre) allows the experimenter to test whether a recorded unit is reliably driven by laser stimulation and thus can be classified as the ChR2-expressing cell type. Right, example recording of a REM-active GAD2-neuron in the ventral medulla along with colour-coded brain states. Data adapted from ref. 74. **c**, Calcium imaging from identified cell types. Top left, using a microendoscope coupled to a head-mountable miniaturized camera, the calcium responses of identified cell types can be imaged in deep brain structures of freely moving mice. Bottom left, GAD2 neurons in the dorsal pons expressing the genetically encoded calcium indicator GCaMP6. Right, calcium ($\Delta F/F$) transients of five wake-active neurons (red circles in bottom left picture) along with colour-coded brain states. Data adapted from ref. 85.

Histaminergic neurons could also inhibit the sleep neurons through their co-release of GABA[42]. In addition, sleep-active POA neurons express κ and μ opioid receptors[43], both of which can activate potassium channels and inhibit voltage-gated calcium channels, thereby reducing the excitability of the neurons. However, while local application of a μ receptor agonist within the VLPO promotes wakefulness, a κ receptor agonist promotes NREM sleep, suggesting the existence of other sites of action besides the sleep-active neurons. Retrograde tracing combined with *in situ* hybridization suggests that neurons in the TMN release endomorphin (a μ receptor agonist), whereas those in the lateral parabrachial nucleus release dynorphin (a κ receptor agonist). Notably, dynorphin is also expressed within the POA, raising the possibility of a local source for promoting sleep.

Whereas inputs from wake-promoting neurons are generally inhibitory, VLPO neurons were recently shown to be excited by physiological concentrations of glucose and might thus mediate the sleep-promoting effect of glucose infusion into the VLPO[44]. This is opposite to the wake-promoting orexin neurons, which are inhibited by high sucrose levels[45]. Thus, the activity of both wake- and sleep-promoting neurons might be sensitive to the energy status of the animal, which may explain why we feel sleepy after eating a meal that is high in sugar. Moreover, the POA contains thermosensitive neurons that are activated by either warming or cooling of the POA within a physiologically relevant temperature range[46]. Notably, the majority of warm-sensitive neurons are also sleep-active[47], which may provide a mechanistic link between the control of sleep and body temperature. An important direction for future studies is to identify additional inputs to the POA sleep-promoting neurons that allow integration of temperature, energy status and other physiological variables for the optimal control of sleep.

The studies summarized above have provided important insights into how the POA contributes to sleep regulation. However, single-unit recordings and c-Fos staining indicate that the sleep-active neurons are not restricted to the VLPO or MnPO, and even within these regions they are spatially intermingled with wake-active neurons, many of which are also GABAergic[48–50]. Although galanin is expressed in many VLPO sleep-active neurons[26,27,32], it also labels the nearby medial preoptic nucleus, which is involved in parental behaviours[51]. A crucial step in dissecting the POA sleep circuit is to identify molecular markers that specifically label sleep-active and sleep-promoting neurons to allow selective manipulation, recording and input and output tracing from these neurons.

Basal forebrain

While the sleep-promoting POA neurons appear to inhibit wake-promoting circuits in multiple regions of the brain, recent work in the basal forebrain (BF, adjacent to the POA) showed that sleep neurons can also suppress wake-promoting neurons locally. Lesion studies have suggested that the BF is important for both sleep and wakefulness[19,31,52], and it contains spatially intermingled sleep- and wake-active neurons[49,53,54].

There are three major cell types in the BF: cholinergic, glutamatergic and GABAergic. Juxtacellular recording and labelling in head-fixed rats followed by immunohistochemical staining showed that the cholinergic neurons are wake- and REM-active[55], and that optogenetic activation of these neurons promotes wakefulness[56,57]. Cell-type-specific channelrhodopsin (ChR2) tagging and optrode recording in freely moving mice showed that glutamatergic and parvalbumin (PV)-expressing GABAergic neurons are also wake- and REM-active, and that activation of these neurons promotes wakefulness. In contrast, a subpopulation of somatostatin (SOM)-expressing GABAergic neurons are NREM-active and activation of this SOM population promotes NREM sleep[57].

The intermingling of multiple cell types in the BF provides ample opportunity for local synaptic interactions, which have been analysed by ultrastructural studies[58], *in vitro* pharmacology[59] and *in vivo* microdialysis[60]. Indeed, ChR2-assisted circuit mapping in BF slices revealed functional synapses for most pairs of cell types[57] (Fig. 4c). In particular, SOM GABAergic neurons provide strong inhibition to cholinergic, glutamatergic, and PV-expressing GABAergic neurons, all of which are wake-promoting. Thus, broad inhibition of multiple wake-promoting cell types, via either local synapses[57] or long-range projections[22,26], appears to be a common feature of sleep-promoting GABAergic neurons[13] (Fig. 4). The local glutamatergic → cholinergic, glutamatergic → PV-expressing, and cholinergic → PV-expressing neuron excitation detected in these experiments also provides useful insights into the inputs that shape the wake- and REM-active properties of these BF neurons and the local circuits that are recruited when a given cell type (for example, glutamatergic) is activated optogenetically.

Lateral hypothalamus

Similar to the POA and BF, the lateral hypothalamus also contains intermingled sleep- and wake-active neurons, with a subset of the sleep-active neurons expressing melanin-concentrating hormone (MCH)[61,62]. A study based on juxtacellular recordings showed that MCH neurons are sleep-active, with maximal firing rates during REM sleep[63]. Brief optogenetic activation of MCH neurons (tens of seconds per trial) increased the transitions from NREM to REM sleep[64] and prolonged the durations of REM sleep episodes[62], indicating that MCH neuron activation enhanced both the initiation and maintenance of REM sleep. However, archaerhodopsin

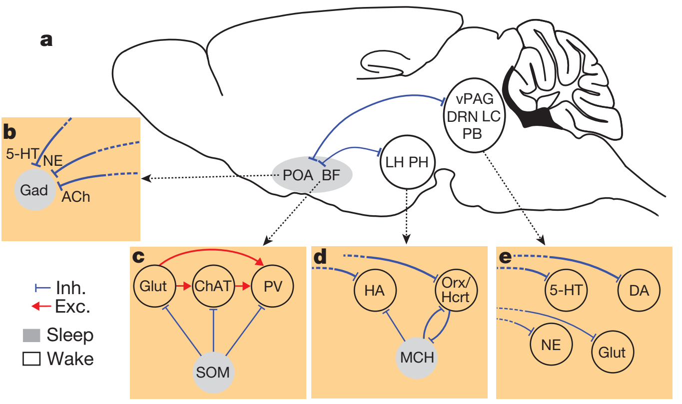


Figure 4 | Circuit diagram for forebrain sleep-promoting mechanisms. **a**, Sleep-active neurons in the POA inhibit wake-active neurons in the lateral and posterior hypothalamus (LH and PH) and in the brainstem including the dorsal raphe nucleus (DRN), locus coeruleus (LC), ventral periaqueductal grey (vPAG), and parabrachial nucleus (PB). **b**, Sleep-active GABAergic (Gad) neurons in the POA are inhibited by norepinephrine (NE), acetylcholine (ACh) and a subgroup by serotonin (5-HT). **c**, Within the basal forebrain, somatostatin (SOM) neurons inhibit neighbouring wake-promoting glutamatergic (Glut), cholinergic (ChAT), and parvalbumin (PV)-expressing GABAergic neurons. Glutamatergic neurons powerfully promote wakefulness through excitation of ChAT and PV neurons. **d**, Sleep-active neurons in the POA inhibit wake-promoting histaminergic (HA) and orexin (also known as hypocretin) (Orx/Hcrt) neurons in the LH and PH. Sleep-active melanin-concentrating hormone (MCH)-expressing neurons suppress HA and Orx/Hcrt neurons. MCH and Orx/Hcrt neurons mutually inhibit each other. **e**, Sleep-active neurons in the POA inhibit wake-promoting neurons in the brainstem including NE-, dopamine (DA)-, 5-HT- and glutamate (Glut)-expressing neurons in the LC, vPAG, DRN, and PB.

(Arch)- or halorhodopsin (NpHR)-mediated silencing of these neurons caused no pronounced change in the amount or duration of REM sleep[62,64], suggesting that MCH neuron activity is sufficient but perhaps not necessary. Notably, selective ablation of MCH neurons using cell-type-specific expression of diphtheria toxin A caused a decrease of NREM sleep without affecting REM sleep[64], suggesting that chronic activity of these neurons is important for NREM sleep. This notion is also supported by the finding that chronic optogenetic activation of MCH neurons (24 h) can enhance NREM as well as REM sleep[65].

The sleep-promoting effect of MCH neurons could be mediated in part by their inhibitory influence on nearby orexin neurons[66], which promote wakefulness[29] and indirectly inhibit MCH neurons[67] (Fig. 4d). Optogenetic activation of MCH neurons also induces inhibitory post-synaptic currents in histaminergic and other neurons in the lateral and posterior hypothalamus, a pathway that is thought to be important for the *in vivo* effect[62]. Notably, these inhibitory responses are primarily caused by GABA rather than MCH synaptic transmission, consistent with the finding that MCH neuron activation can promote REM sleep even in the absence of MCH receptors[62].

Control of NREM versus REM sleep

After suppression of wakefulness by forebrain sleep neurons, the brain alternates between NREM and REM sleep. The duration of the so-called ultradian NREM/REM cycle varies across species; 90–120 min in humans (Fig. 1b) and 10–20 min in rodents (Fig. 1d, e). Following the discovery of REM sleep and its associated dreaming in the 1950s[2,68], the neural mechanisms controlling this brain state have been under active investigation.

By making surgical transections at various rostrocaudal levels of the cat brain and measuring the neural signatures of REM sleep on each side of the cut, Jouvet concluded that the brainstem is both necessary and sufficient for REM sleep generation[68]. Two prominent neuromodulatory systems show opposite firing rate changes at NREM to REM transitions: monoaminergic neurons cease to fire (REM off) and cholinergic neurons become highly active (REM on)[36]. This led to the formulation of a model for the ultradian cycle based on the reciprocal interactions between these two neuronal populations[36,69]. Subsequent studies, however, have pointed to more prominent roles of glutamatergic and GABAergic neurons in REM sleep generation[70–74] (Fig. 5a). Furthermore, recent studies have identified several groups of glutamatergic and GABAergic brainstem neurons that specifically promote NREM sleep[71,74–76] (Fig. 5b). The antagonistic interactions between these REM- and NREM-promoting neurons within the brainstem may play important roles in controlling the ultradian cycle.

REM-promoting neurons in the brainstem

Following Jouvet's landmark transection studies[68], lesion and pharmacological experiments have consistently identified the dorsolateral pons as an important region for REM sleep generation[5,72,77,78]. This region contains diverse cell types, including cholinergic, noradrenergic, glutamatergic and GABAergic neurons, and their respective functions in REM sleep control are still under debate[5,36,70,72].

Cholinergic neurons in the pedunculopontine tegmentum (PPT) and laterodorsal tegmentum (LDT) are wake- and REM-active neurons[79]. Early pharmacological experiments showed a powerful effect of cholinergic agonists in REM sleep generation, but application of antagonists or lesions of cholinergic neurons have yielded variable results[5,72,80,81]. In a recent study, the role of pontine cholinergic neurons was tested using optogenetics[82]. Activation of these neurons during NREM sleep increased the frequency but not the duration of REM sleep episodes, suggesting that these neurons contribute to REM sleep initiation but perhaps not to maintenance.

Extracellular recordings in the cat suggested that there are also non-cholinergic REM-on neurons in the dorsal pons[83]. Using c-Fos immunohistochemistry to detect REM-active cells, some investigators identified glutamatergic neurons in the sublateral dorsal nucleus (SLD)[70,72], while others also found GABAergic neurons in the region[72,84]. Juxtacellular recordings in head-fixed rats confirmed the existence of REM-active GABAergic neurons[79], and cell-type-specific calcium imaging in freely moving mice revealed REM-active glutamatergic neurons[85]. Pharmacological activation of the SLD induced a REM sleep-like state with EEG desynchronization and muscle atonia[77], and conditional knockout of *Vglut2* (the gene encoding vesicular glutamate transporter 2) caused fragmentation of REM sleep and reduced its amount[86], indicating an important role of SLD glutamatergic neurons in REM sleep generation.

In addition to the pons, c-Fos immunohistochemistry showed that the medulla also contains REM-active neurons[73]. While previous studies have emphasized the function of the ventral medulla in generating the muscle atonia associated with REM sleep through its projections to the spinal cord[87,88], a recent study demonstrated that rostrally projecting ventral medulla neurons are critical in generating REM sleep[74]. Optogenetic activation of GABAergic neurons in the ventral medulla or their axons projecting to the midbrain induced a marked increase in the probability of NREM to REM transitions, but activating glutamatergic neurons in the same region reliably induced wakefulness, attesting to the importance of cell-type-specific manipulation of neuronal activity. Using a closed-loop stimulation protocol afforded by rapid optogenetic control of neuronal activity (Fig. 2c), activation or silencing of the GABAergic neurons was found to respectively prolong or shorten the duration of each REM sleep episode, indicating that the activity of these neurons is also important for maintaining REM sleep. Furthermore, optrode recording from ChR2-tagged GABAergic neurons shows that their firing rates increase gradually over a period of tens of seconds before the onset of REM sleep and are sustained at a high level during each REM sleep episode, a temporal profile well suited for their roles in the induction and maintenance of REM sleep (Fig. 3b). Importantly, their natural firing rates during REM sleep are higher than the laser stimulation frequency necessary to induce REM sleep, indicating that their endogenous activity is sufficient for the REM-promoting effects demonstrated with optogenetic manipulations. In

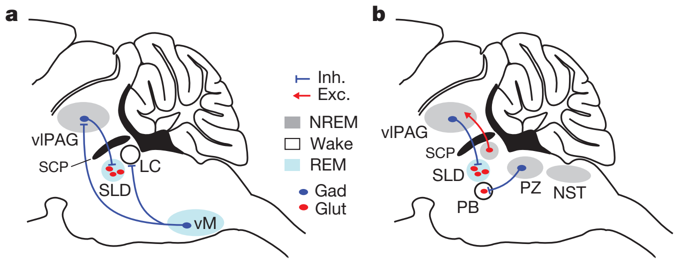


Figure 5 | Brainstem circuits controlling REM and NREM sleep. **a**, Brainstem circuit promoting REM sleep, in which neuronal interactions characterized in recent studies are selectively highlighted. Glutamatergic (Glut) REM-promoting neurons in the SLD are probably inhibited by NREM-promoting GABAergic (Gad) vlPAG neurons. Activation of GABAergic ventral medulla (vM) neurons that innervate the vlPAG might thus disinhibit the SLD and therefore promote NREM to REM transitions. Neurons in the ventral and dorsal medulla probably inhibit noradrenergic neurons in the LC to maintain REM sleep and delay awakening. **b**, NREM-promoting circuit in the brainstem. GABAergic neurons in the vlPAG, parafacial zone (PZ) and glutamatergic neurons located ventromedial to the superior cerebellar peduncle (SCP) promote NREM sleep and strongly suppress both wakefulness and REM sleep. GABAergic vlPAG neurons, probably excited by the glutamatergic NREM-promoting neurons close to the SCP, inhibit the SLD and thus suppress REM sleep. Inhibition of the nearby medial parabrachial nucleus (PB) by GABAergic PZ neurons suppresses wakefulness and promotes NREM sleep. Electrical stimulation of the nucleus of the solitary tract (NTS) also increases sleep, but the underlying cell types and synaptic interactions are unknown.

addition to spike rate, such cell-type-specific recordings *in vivo* may also reveal brain-state-dependent changes in firing pattern (for example, burst versus tonic), which may strongly influence the release of neuropeptides important for regulating brain states[89].

NREM-promoting neurons in the brainstem

Early transection and pharmacological inactivation experiments pointed to a synchronizing mechanism in the medulla, which reduces the magnitude and duration of EEG desynchronization induced by stimulating the ascending reticular activating system[90]. Subsequent studies showed that electrical stimulation of the nucleus of the solitary tract (NTS) synchronized the EEG and increased sleep[91]. The finding of NREM-active neurons in the NTS[92] further suggests the physiological relevance of this region, although the specific cell type that mediates the synchronizing effect remains unknown.

Recent studies have identified additional cell groups in the midbrain, pons and medulla that promote NREM sleep (Fig. 5b). Pharmacogenetic activation of a population of glutamatergic neurons in the dorsolateral pons, located ventromedial to the superior cerebellar peduncle (SCP), enhanced NREM sleep[71]. c-Fos immunohistochemistry showed that many GABAergic neurons in the parafacial zone (PZ, located in the rostral medulla lateral and dorsal to the facial nucleus) are sleep-active[76]. Lesion of PZ neurons or deletion of *Vgat* (the gene encoding vesicular GABA/glycine transporter) in these neurons strongly increased wakefulness, whereas pharmacogenetic activation of VGAT-expressing PZ neurons increased NREM sleep[75,76]. This effect is thought to be mediated by inhibition of glutamatergic neurons in the medial parabrachial nucleus, which are known to be important for wakefulness and arousal[31]. Furthermore, optogenetic[74] or pharmacogenetic[71] activation of GABAergic neurons in the ventrolateral periaqueductal grey (vlPAG) or the adjacent deep mesencephalic reticular nucleus (DpMe) substantially increased NREM sleep, thus revealing yet another brainstem neuronal population promoting NREM sleep.

NREM/REM antagonism

Note that the hypothalamic and BF sleep neurons seem to promote NREM and/or REM sleep primarily by suppressing wakefulness (Fig. 4). In contrast, the brainstem sleep neurons discussed above also contribute to NREM/REM antagonism. For instance, pharmacogenetic activation of PZ GABAergic neurons[75] or the glutamatergic neurons ventromedial to the SCP[71] leads to a strong suppression of REM sleep whilst enhancing NREM sleep. Optogenetic[74] or pharmacogenetic[71] activation of vlPAG or DpMe GABAergic neurons also suppresses REM sleep in addition to wakefulness, consistent with previous lesion[72] and pharmacological inactivation[73,93] experiments. Conversely, activation of the ventral medulla GABAergic neurons promotes REM sleep by greatly enhancing NREM to REM transitions[74], thus effectively suppressing NREM sleep.

The circuit basis for the antagonistic relationship between NREM and REM sleep probably involves mutual inhibition between the NREM- and REM-promoting neurons. Gating of REM sleep by the vlPAG and DpMe could be mediated by their GABAergic innervation of the SLD[72], while trans-synaptic retrograde tracing with a modified rabies virus[11] revealed that vlPAG GABAergic neurons were directly inhibited by the ventral medulla[74]. Optogenetic activation of GABAergic axons of ventral medulla neurons within the vlPAG was sufficient to trigger and maintain REM sleep, suggesting that the inhibition of the vlPAG by the ventral medulla is an important mechanism promoting NREM to REM transitions. These inhibitory interactions between the NREM- and REM-promoting neurons might form the core of the ultradian oscillator. Notably, the duration of the ultradian cycle appears to be correlated with brain size[3]. In future studies it would be important to understand the temporal dynamics of the neuronal interactions that determine the duration of NREM/REM cycles, and how these dynamic properties are related to brain size.

Besides the NREM-promoting neurons, REM neurons must also interact with wake-promoting neurons. For example, GABAergic neurons in the ventral[74,94] and dorsal[95] medulla probably inhibit the wake-active, REM-off noradrenergic neurons in the LC, which may be important for the maintenance of REM sleep by delaying awakening. Spontaneous awakening in humans is most likely to occur at the end of a REM sleep episode, and rodents typically wake up after REM sleep rather than immediately transitioning back into NREM sleep. What processes favour REM to wake over REM to NREM transitions? Notably, although some wake-promoting neurons are silent during REM sleep (for example, monoaminergic and orexinergic neurons), the majority of neurons within the dorsolateral pons are both wake- and REM-active[79,85]. In the basal forebrain, glutamatergic, cholinergic, and PV-expressing GABAergic neurons are also REM- and wake-active, but their activation promotes wakefulness[57]. Such activation of wake-promoting neurons during REM sleep might bias the transition into wake rather than NREM sleep at the end of each REM episode.

Homeostatic and circadian regulation of sleep

While the rapid transitions between wake, NREM and REM states are controlled by mutual inhibitory interactions among the neuronal groups promoting these states, sleep is also known to be regulated by homeostatic and circadian processes on much slower timescales[96]. How these processes influence the sleep–wake network is only partially understood.

Sleep pressure and homeostasis

Homeostatic regulation refers to the fact that after prolonged wakefulness the animal tends to sleep for longer periods and/or at higher intensities. The increased sleep intensity is reflected by increased slow-wave activity (SWA, between 0.5 and 4.5 Hz) in the EEG, which decays gradually during the course of recovery sleep[97]. SWA thus serves as an excellent marker for sleep pressure. The homeostatic regulation of sleep is under genetic control[98]; a variety of genes have been shown to affect sleep homeostasis, some of which are also involved in the circadian regulation of sleep[16].

More than a century ago, Ishimori and Pieron found that injection of the cerebrospinal fluid from a sleep-deprived dog into a normal one triggered sleep[99,100], leading to the idea that sleep pressure exerts its impact through chemical factors (somnogens) that accumulate during wakefulness. In particular, adenosine has been studied extensively as a somnogen. The extracellular concentration of adenosine increases with the time spent awake and declines during recovery sleep in both the basal

forebrain[101] and cortex[102]. One source of adenosine could be the BF, with its cholinergic neurons playing a particularly important role[103]. There is also strong evidence for the involvement of astrocytes in regulating the adenosine concentration[104,105].

Adenosine regulates neuronal activity via two major classes of adenosine receptors: the inhibitory A1 receptors that are distributed throughout the brain and the excitatory A2A receptors that are mainly localized in the striatum, nucleus accumbens and the olfactory tubercle. In A2A receptor knockout mice, caffeine (an A1 and A2A receptor antagonist) failed to promote wakefulness[106] and deprivation-induced NREM sleep rebound was reduced[107]. *In vivo* application of an A2A receptor agonist increased NREM sleep and enhanced c-Fos expression in the VLPO and MnPO[108,109], whereas application of an antagonist in the VLPO attenuated sleep-deprivation-induced increases in the firing rates of sleep-active neurons[110]. In brain slices, VLPO neurons were shown to be activated directly or indirectly via A2A receptors[111,112]. By contrast, A1 receptors mediate the inhibitory effect of adenosine on wake-active neurons *in vitro*, including basal forebrain cholinergic neurons[113] and hypothalamic orexin neurons[114]. *In vivo* application of adenosine or A1 receptor agonists also increased NREM sleep[115]. Although A1 receptor knockout mice showed no change in the homeostatic regulation of the amount of sleep[116], conditional deletion of A1 receptors in the forebrain and brainstem attenuated the rebound in SWA induced by sleep deprivation[104,117]. Thus, adenosine seems to promote sleep by simultaneously activating sleep neurons through A2A receptors and suppressing wake neurons through A1 receptors.

Besides adenosine, other somnogenic factors have been identified, including prostaglandin D2, nitric oxide, growth hormone releasing hormone and cytokines (for review see ref. 118). In particular, prostaglandin D2 has been shown to be a powerful somnogen. Injection of prostaglandin D2 into the POA or the nearby subarachnoid space increases sleep and c-Fos expression in the VLPO[119], an effect that is probably mediated via the activation of prostaglandin receptors that in turn increase adenosine levels[120].

In addition to acting on the hypothalamic and basal forebrain neurons that control global brain states, sleep homeostasis has a strong local component. Cortical regions that have been more active during the preceding wake period exhibit stronger SWA during sleep[121,122], which requires a local mechanism for measuring recent activity and synchronizing the cortical population. Recently, a class of cortical interneurons expressing neuronal nitric oxide synthase (nNOS) has been proposed as a link between sleep pressure and cortical SWA[123,124]. Expression of c-Fos in nNOS neurons correlates with SWA and as these neurons have long-range intracortical projections, nNOS neurons might therefore be well suited for synchronizing the activity of neural populations[125].

While most of the studies on sleep pressure have focused on NREM sleep, REM sleep is also under strong homeostatic control[126], which is probably separate from NREM sleep homeostasis[127]. The molecular and circuit mechanisms underlying REM sleep homeostasis represent an important frontier that remains largely unexplored.

Circadian rhythm of sleep

Circadian modulation of sleep depends critically on the suprachiasmatic nucleus (SCN) in the hypothalamus, the master pacemaker of the whole organism. Lesion of the SCN or its downstream target regions eliminates the daily rhythm of sleep without markedly affecting its amount[128,129], suggesting that the SCN is not part of the core circuit for sleep generation, but it regulates the circadian timing of sleep.

SCN neuron spiking was shown to play a key role in regulating both the molecular clock and behavioural rhythm, as optogenetic activation or suppression of SCN activity was sufficient to alter the phase and periodicity of clock gene expression and of the sleep–wake cycle[130]. The firing rates of SCN neurons are high during the subjective day and low during the night, regardless of whether the animal is diurnal or nocturnal. Such circadian variation of electrical activity is controlled by both the molecular clock driven by multiple transcriptional/translational feedback loops[131] that regulates the intrinsic excitability of SCN neurons[132] and the synaptic inputs signalling the light–dark cycle of the environment[133]. Notably, *in vivo* multiunit recordings showed that, superimposed on the slow circadian variation, SCN neuron firing rates also change with the sleep–wake states on a timescale of seconds[134]. This is probably caused by synaptic inputs from neurons involved in sleep–wake regulation, such as cholinergic and monoaminergic neurons[135,136]. Given such ultradian firing rate modulations, it would be interesting to know whether SCN activity can exert rapid influences on brain states in the order of seconds to minutes, in addition to its well-known circadian effect in the order of hours. In addition, the SCN consists of multiple cell types including vasoactive intestinal peptide-positive and arginine vasopressin-positive cells. Whether different types of SCN neurons play distinct roles in sleep–wake regulation remains to be investigated.

The SCN projects to multiple target regions to coordinate a variety of physiological functions. Among these targets the dorsomedial hypothalamic nucleus (DMH) may play a particularly important role in sleep–wake regulation. Lesion of the DMH largely eliminated the sleep–wake circadian rhythm[128]. In addition, a study using a pseudo-rabies virus for trans-synaptic retrograde tracing showed that the DMH provides an important relay from the SCN to the LC[137], which contains wake-promoting noradrenergic neurons[138]. Importantly, the DMH comprises both glutamatergic and GABAergic neurons that appear to innervate both sleep- and wake-promoting circuits[128]. Understanding the functional organization of this structure and how it mediates the circadian modulation of sleep again requires cell-type-specific recording, manipulation and circuit mapping.

Looking forward

The past few years have witnessed rapid progress in our understanding of the neural circuits controlling sleep, largely enabled by technical innovations that allow measurement and manipulation of neuronal activity from genetically defined cell types and tracing of their synaptic inputs and outputs. These new experimental approaches have led to the identification of additional neuronal populations in the hypothalamus, BF and brainstem that promote NREM and REM sleep, and their local and long-range connections are beginning to be delineated.

Of course, while these novel techniques have opened new avenues for investigation, it is always important to keep in mind their limitations. For example, while optogenetics provides an easily applicable method for controlling neural activity with cell-type specificity and high temporal precision (Fig. 2), the results should be interpreted with care. In sleep–wake control circuits many neurons co-release neuropeptides and other modulators together with GABA or glutamate, but the release mechanisms show differential dependence on the rate and temporal pattern (for example, burst versus tonic) of spiking[89]. To assess whether the observed effects of optogenetic manipulations are physiological, it is important to measure the natural activity of the neurons across different behavioural states[74,79,139]. Combining optogenetic manipulations with microdialysis[60] or with blockade of neurotransmitter/modulator receptors using genetic or pharmacological approaches[62] will also provide important insights into the downstream signalling pathways mediating the behavioural effects.

One of the most exciting developments in the past few years is the identification of new sleep-promoting neurons, and the growth of this list is likely to continue. An important question is whether these sleep-promoting cell groups are organized hierarchically, in recurrent loops, or whether they work in parallel to control different aspects of sleep. NREM and REM sleep alternate in an ultradian rhythm. Although we know a great deal about how transcriptional/translational feedback loops generate the circadian rhythm and how neuronal biophysical/synaptic properties underlie network oscillations on a millisecond–to–second timescale, we know very little about how the brain generates the ultradian REM–NREM alternation on a minute–to–hour time scale. The control of REM–NREM alternations probably depends on both the synaptic interactions among spiking neurons and slower translational/transcriptional processes and

accumulation of chemical substances. Uncovering these circuit and molecular mechanisms underlying the ultradian oscillation is not only important for sleep research, but will also bridge an important gap in our understanding of brain rhythms in general.

In addition to circadian and homeostatic regulations, sleep is strongly influenced by a variety of emotional and physiological parameters, such as stress, pain, hunger and body temperature. Some of the interactions between sleep and these other processes may be mediated by common neurons shared between different control circuits. For example, while MCH neuron activity contributes to sleep regulation[62,64,65], the neuropeptide MCH has been implicated in feeding[140]; and while activation of ventral medulla GABAergic neurons during NREM sleep reliably induces REM sleep, it enhances eating during wakefulness[74]. Thus both hypothalamic MCH neurons and ventral medulla GABAergic neurons may help to link the regulation of sleep and feeding. The regulation of sleep by the emotional, thermal and nutritional state of the animal probably also involves synaptic inputs from these circuits to the sleep–wake control network. Furthermore, both sleep- and wake-promoting neurons are probably modulated by humoral factors such as stress hormones, cytokines and glucose. Identification of the corresponding receptors expressed in sleep and wake neurons will allow us to uncover the molecular basis for the interactions between the circuits controlling sleep and other biological functions that are essential for survival.

Received 19 May 2015; accepted 17 August 2016.

1. Shein-Idelson, M., Ondracek, J. M., Liaw, H.-P., Reiter, S. & Laurent, G. Slow waves, sharp waves, ripples, and REM in sleeping dragons. *Science* **352,** 590–595 (2016).
This study reports for the first time the existence of REM- and NREM-like sleep stages in a reptile, the Australian dragon *Pogona vitticeps*. Comparative analysis might shed light on the common circuitry underlying ultradian rhythms in reptiles and mammals.
2. Aserinsky, E. & Kleitman, N. Regularly occurring periods of eye motility, and concomitant phenomena, during sleep. *Science* **118,** 273–274 (1953).
3. Zepelin, H., Siegel, J. M. & Tobler, I. in *Principles and practice of sleep medicine* **4,** 91–100 (eds Kryger, M. H. *et al*., Elsevier Saunders, 2005).
4. Moruzzi, G. & Magoun, H. W. Brain stem reticular formation and activation of the EEG. *Electroencephalogr. Clin. Neurophysiol.* **1,** 455–473 (1949).
5. Brown, R. E., Basheer, R., McKenna, J. T., Strecker, R. E. & McCarley, R. W. Control of sleep and wakefulness. *Physiol. Rev.* **92,** 1087–1187 (2012).
6. Saper, C. B., Fuller, P. M., Pedersen, N. P., Lu, J. & Scammell, T. E. Sleep state switching. *Neuron* **68,** 1023–1042 (2010).
7. Deisseroth, K. Optogenetics. *Nat. Methods* **8,** 26–29 (2011).
8. Armbruster, B. N., Li, X., Pausch, M. H., Herlitze, S. & Roth, B. L. Evolving the lock to fit the key to create a family of G protein-coupled receptors potently activated by an inert ligand. *Proc. Natl Acad. Sci. USA* **104,** 5163–5168 (2007).
9. Tian, L., Hires, S. A. & Looger, L. L. Imaging neuronal activity with genetically encoded calcium indicators. *Cold Spring Harb. Protoc.* **2012,** 647–656 (2012).
10. Oh, S. W. *et al.* A mesoscale connectome of the mouse brain. *Nature* **508,** 207–214 (2014).
11. Osakada, F. & Callaway, E. M. Design and generation of recombinant rabies virus vectors. *Nat. Protocols* **8,** 1583–1601 (2013).
12. Rihel, J. & Schier, A. F. Sites of action of sleep and wake drugs: insights from model organisms. *Curr. Opin. Neurobiol.* **23,** 831–840 (2013).
13. Sehgal, A. & Mignot, E. Genetics of sleep and sleep disorders. *Cell* **146,** 194–207 (2011).
14. Tononi, G. & Cirelli, C. Sleep and the price of plasticity: from synaptic and cellular homeostasis to memory consolidation and integration. *Neuron* **81,** 12–34 (2014).
15. Trojanowski, N. F. & Raizen, D. M. Call it worm sleep. *Trends Neurosci*. **39,** 54–62 (2016).
16. Andretic, R., Franken, P. & Tafti, M. Genetics of sleep. *Annu. Rev. Genet.* **42,** 361–388 (2008).
17. Von Economo, C. Sleep as a problem of localization. *J. Nerv. Ment. Dis.* **71,** 249–259 (1930).
18. Nauta, W. J. Hypothalamic regulation of sleep in rats; an experimental study. *J. Neurophysiol.* **9,** 285–316 (1946).
19. McGinty, D. J. & Sterman, M. B. Sleep suppression after basal forebrain lesions in the cat. *Science* **160,** 1253–1255 (1968).
20. Sallanon, M. *et al.* Long-lasting insomnia induced by preoptic neuron lesions and its transient reversal by muscimol injection into the posterior hypothalamus in the cat. *Neuroscience* **32,** 669–683 (1989).
21. Lin, J.-S., Sakai, K., Vanni-Mercier, G. & Jouvet, M. A critical role of the posterior hypothalamus in the mechanisms of wakefulness determined by microinjection of muscimol in freely moving cats. *Brain Res*. **479,** 225–240 (1989).
22. Sherin, J. E., Shiromani, P. J., McCarley, R. W. & Saper, C. B. Activation of ventrolateral preoptic neurons during sleep. *Science* **271,** 216–219 (1996).
23. Gong, H. *et al.* Activation of c-fos in GABAergic neurones in the preoptic area during sleep and in response to sleep deprivation. *J. Physiol. (Lond.)* **556,** 935–946 (2004).
24. Lu, J., Greco, M. A., Shiromani, P. & Saper, C. B. Effect of lesions of the ventrolateral preoptic nucleus on NREM and REM sleep. *J. Neurosci.* **20,** 3830–3842 (2000).
25. Zhang, Z. *et al.* Neuronal ensembles sufficient for recovery sleep and the sedative actions of $\alpha 2$ adrenergic agonists. *Nat. Neurosci.* **18,** 553–561 (2015).
This study used a pharmacogenetic approach to activate POA neurons that expressed c-Fos following sustained sleep to show that they indeed promote sleep.
26. Sherin, J. E., Elmquist, J. K., Torrealba, F. & Saper, C. B. Innervation of histaminergic tuberomammillary neurons by GABAergic and galaninergic neurons in the ventrolateral preoptic nucleus of the rat. *J. Neurosci.* **18,** 4705–4721 (1998).
27. Steininger, T. L., Gong, H., McGinty, D. & Szymusiak, R. Subregional organization of preoptic area/anterior hypothalamic projections to arousal-related monoaminergic cell groups. *J. Comp. Neurol.* **429,** 638–653 (2001).
28. Yoshida, K., McCormack, S., España, R. A., Crocker, A. & Scammell, T. E. Afferents to the orexin neurons of the rat brain. *J. Comp. Neurol.* **494,** 845–861 (2006).
29. Adamantidis, A. R., Zhang, F., Aravanis, A. M., Deisseroth, K. & de Lecea, L. Neural substrates of awakening probed with optogenetic control of hypocretin neurons. *Nature* **450,** 420–424 (2007).
This study was the first to apply optogenetics to investigate circuits involved in sleep–wake regulation and it provided causal evidence that orexin neurons promote wakefulness.
30. Lu, J., Jhou, T. C. & Saper, C. B. Identification of wake-active dopaminergic neurons in the ventral periaqueductal gray matter. *J. Neurosci.* **26,** 193–202 (2006).
31. Fuller, P. M., Sherman, D., Pedersen, N. P., Saper, C. B. & Lu, J. Reassessment of the structural basis of the ascending arousal system. *J. Comp. Neurol.* **519,** 933–956 (2011).
32. Gaus, S. E., Strecker, R. E., Tate, B. A., Parker, R. A. & Saper, C. B. Ventrolateral preoptic nucleus contains sleep-active, galaninergic neurons in multiple mammalian species. *Neuroscience* **115,** 285–294 (2002).
33. Steininger, T. L., Alam, M. N., Gong, H., Szymusiak, R. & McGinty, D. Sleep-waking discharge of neurons in the posterior lateral hypothalamus of the albino rat. *Brain Res*. **840,** 138–147 (1999).
34. Nitz, D. & Siegel, J. M. GABA release in the locus coeruleus as a function of sleep/wake state. *Neuroscience* **78,** 795–801 (1997).
35. Nitz, D. & Siegel, J. GABA release in the dorsal raphe nucleus: role in the control of REM sleep. *Am. J. Physiol.* **273,** R451–R455 (1997).
36. Hobson, J. A., McCarley, R. W. & Wyzinski, P. W. Sleep cycle oscillation: reciprocal discharge by two brainstem neuronal groups. *Science* **189,** 55–58 (1975).
37. Schönrock, B., Büsselberg, D. & Haas, H. L. Properties of tuberomammillary histamine neurones and their response to galanin. *Agents Actions* **33,** 135–137 (1991).
38. Pieribone, V. A. *et al.* Galanin induces a hyperpolarization of norepinephrine-containing locus coeruleus neurons in the brainstem slice. *Neuroscience* **64,** 861–874 (1995).
39. Chou, T. C. *et al.* Afferents to the ventrolateral preoptic nucleus. *J. Neurosci.* **22,** 977–990 (2002).
40. Gallopin, T. *et al.* Identification of sleep-promoting neurons *in vitro*. *Nature* **404,** 992–995 (2000).
41. Liu, Y.-W., Li, J. & Ye, J.-H. Histamine regulates activities of neurons in the ventrolateral preoptic nucleus. *J. Physiol. (Lond.)* **588,** 4103–4116 (2010).
42. Yu, X. *et al.* Wakefulness is governed by GABA and histamine cotransmission. *Neuron* **87,** 164–178 (2015).
This study demonstrated that histaminergic neurons in the TMN co-release histamine and GABA. Abolishing VGAT expression in histaminergic neurons increases wakefulness. Hence, the wake-promoting effect of histamine might be balanced by the co-release of GABA.
43. Greco, M.-A. *et al.* Opioidergic projections to sleep-active neurons in the ventrolateral preoptic nucleus. *Brain Res*. **1245,** 96–107 (2008).
44. Varin, C. *et al.* Glucose induces slow-wave sleep by exciting the sleep-promoting neurons in the ventrolateral preoptic nucleus: a new link between sleep and metabolism. *J. Neurosci.* **35,** 9900–9911 (2015).
This study demonstrated that sleep-active neurons in the preoptic area are excited by glucose, which may provide a neuronal link between metabolism and sleep.
45. Yamanaka, A. *et al.* Hypothalamic orexin neurons regulate arousal according to energy balance in mice. *Neuron* **38,** 701–713 (2003).
46. Boulant, J. A. & Dean, J. B. Temperature receptors in the central nervous system. *Annu. Rev. Physiol.* **48,** 639–654 (1986).
47. Alam, M. N., McGinty, D. & Szymusiak, R. Neuronal discharge of preoptic/anterior hypothalamic thermosensitive neurons: relation to NREM sleep. *Am. J. Physiol.* **269,** R1240–R1249 (1995).
48. Szymusiak, R., Alam, N., Steininger, T. L. & McGinty, D. Sleep-waking discharge patterns of ventrolateral preoptic/anterior hypothalamic neurons in rats. *Brain Res*. **803,** 178–188 (1998).

49. Takahashi, K., Lin, J.-S. & Sakai, K. Characterization and mapping of sleep-waking specific neurons in the basal forebrain and preoptic hypothalamus in mice. *Neuroscience* **161,** 269–292 (2009).
A tour de force study recording hundreds of neurons in the preoptic area and basal forebrain during the sleep–wake cycle, demonstrating a staggering functional diversity among the neurons.
50. Modirrousta, M., Mainville, L. & Jones, B. E. GABAergic neurons with α2-adrenergic receptors in basal forebrain and preoptic area express c-Fos during sleep. *Neuroscience* **129,** 803–810 (2004).
51. Wu, Z., Autry, A. E., Bergan, J. F., Watabe-Uchida, M. & Dulac, C. G. Galanin neurons in the medial preoptic area govern parental behaviour. *Nature* **509,** 325–330 (2014).
52. Buzsaki, G. *et al.* Nucleus basalis and thalamic control of neocortical activity in the freely moving rat. *J. Neurosci.* **8,** 4007–4026 (1988).
53. Hassani, O. K., Lee, M. G., Henny, P. & Jones, B. E. Discharge profiles of identified GABAergic in comparison to cholinergic and putative glutamatergic basal forebrain neurons across the sleep-wake cycle. *J. Neurosci.* **29,** 11828–11840 (2009).
54. Szymusiak, R. & McGinty, D. Sleep-waking discharge of basal forebrain projection neurons in cats. *Brain Res. Bull.* **22,** 423–430 (1989).
55. Lee, M. G., Hassani, O. K., Alonso, A. & Jones, B. E. Cholinergic basal forebrain neurons burst with theta during waking and paradoxical sleep. *J. Neurosci.* **25,** 4365–4369 (2005).
This study recorded for the first time from identified cholinergic basal forebrain neurons using the juxtacellular labelling technique. These neurons showed highest activity during waking and REM sleep, when hippocampal theta activity is high.
56. Han, Y. *et al.* Selective activation of cholinergic basal forebrain neurons induces immediate sleep-wake transitions. *Curr. Biol.* **24,** 693–698 (2014).
57. Xu, M. *et al.* Basal forebrain circuit for sleep-wake control. *Nat. Neurosci.* **18,** 1641–1647 (2015).
Combining optogenetic activation, optrode recordings and *in vitro* physiology, this study thoroughly examined how the local interaction of different cell types in the basal forebrain underlies sleep–wake regulation.
58. Zaborszky, L. & Duque, A. Local synaptic connections of basal forebrain neurons. *Behav. Brain Res.* **115,** 143–158 (2000).
59. Yang, C. *et al.* Cholinergic neurons excite cortically projecting basal forebrain GABAergic neurons. *J. Neurosci.* **34,** 2832–2844 (2014).
60. Zant, J. C. *et al.* Cholinergic neurons in the basal forebrain promote wakefulness by actions on neighboring non-cholinergic neurons: an opto-dialysis study. *J. Neurosci.* **36,** 2057–2067 (2016).
This study used a novel opto-dialysis probe (combining optogenetic activation with microdyalisis) to demonstrate that local release of acethylcholine within the basal forebrain is crucial for cortical activation.
61. Hassani, O. K., Henny, P., Lee, M. G. & Jones, B. E. GABAergic neurons intermingled with orexin and MCH neurons in the lateral hypothalamus discharge maximally during sleep. *Eur. J. Neurosci.* **32,** 448–457 (2010).
62. Jego, S. *et al.* Optogenetic identification of a rapid eye movement sleep modulatory circuit in the hypothalamus. *Nat. Neurosci.* **16,** 1637–1643 (2013).
This study applied an optogenetic, closed-loop stimulation protocol to describe the role of hypothalamic MCH neurons in REM sleep maintenance in addition to induction.
63. Hassani, O. K., Lee, M. G. & Jones, B. E. Melanin-concentrating hormone neurons discharge in a reciprocal manner to orexin neurons across the sleep-wake cycle. *Proc. Natl Acad. Sci. USA* **106,** 2418–2422 (2009).
64. Tsunematsu, T. *et al.* Optogenetic manipulation of activity and temporally controlled cell-specific ablation reveal a role for MCH neurons in sleep/wake regulation. *J. Neurosci.* **34,** 6896–6909 (2014).
65. Konadhode, R. R. *et al.* Optogenetic stimulation of MCH neurons increases sleep. *J. Neurosci.* **33,** 10257–10263 (2013).
66. Rao, Y. *et al.* Regulation of synaptic efficacy in hypocretin/orexin-containing neurons by melanin concentrating hormone in the lateral hypothalamus. *J. Neurosci.* **28,** 9101–9110 (2008).
67. Apergis-Schoute, J. *et al.* Optogenetic evidence for inhibitory signaling from orexin to MCH neurons via local microcircuits. *J. Neurosci.* **35,** 5435–5441 (2015).
68. Jouvet, M. Recherches sur les structures nerveuses et les mécanismes responsables des différentes phases du sommeil physiologique. *Arch. Ital. Biol.* **100,** 125–206 (1962).
69. McCarley, R. W. & Hobson, J. A. Neuronal excitability modulation over the sleep cycle: a structural and mathematical model. *Science* **189,** 58–60 (1975).
70. Clément, O., Sapin, E., Bérod, A., Fort, P. & Luppi, P.-H. Evidence that neurons of the sublaterodorsal tegmental nucleus triggering paradoxical (REM) sleep are glutamatergic. *Sleep* **34,** 419–423 (2011).
71. Hayashi, Y. *et al.* Cells of a common developmental origin regulate REM/non-REM sleep and wakefulness in mice. *Science* **350,** 957–961 (2015).
This study identified two sets of glutamatergic neurons in the pontine brainstem of a common developmental lineage that promote NREM sleep and wakefulness. The NREM-promoting neurons might suppress REM sleep by activating GABAergic vlPAG neurons.
72. Lu, J., Sherman, D., Devor, M. & Saper, C. B. A putative flip-flop switch for control of REM sleep. *Nature* **441,** 589–594 (2006).
73. Sapin, E. *et al.* Localization of the brainstem GABAergic neurons controlling paradoxical (REM) sleep. *PLoS One* **4,** e4272 (2009).
74. Weber, F. *et al.* Control of REM sleep by ventral medulla GABAergic neurons. *Nature* **526,** 435–438 (2015).
This study identified a brainstem circuit controlling the induction and maintenance of REM sleep using optogenetics, *in vivo* physiology and viral tracing techniques.
75. Anaclet, C. *et al.* The GABAergic parafacial zone is a medullary slow wave sleep-promoting center. *Nat. Neurosci.* **17,** 1217–1224 (2014).
Identification of a novel group of sleep-promoting neurons in the rostral medulla.
76. Anaclet, C. *et al.* Identification and characterization of a sleep-active cell group in the rostral medullary brainstem. *J. Neurosci.* **32,** 17970–17976 (2012).
77. Boissard, R. *et al.* The rat ponto-medullary network responsible for paradoxical sleep onset and maintenance: a combined microinjection and functional neuroanatomical study. *Eur. J. Neurosci.* **16,** 1959–1973 (2002).
78. George, R., Haslett, W. L. & Jenden, D. J. A cholinergic mechanism in the brainstem reticular formation: induction of paradoxical sleep. *Int. J. Neuropharmacol.* **3,** 541–552 (1964).
79. Boucetta, S., Cissé, Y., Mainville, L., Morales, M. & Jones, B. E. Discharge profiles across the sleep-waking cycle of identified cholinergic, GABAergic, and glutamatergic neurons in the pontomesencephalic tegmentum of the rat. *J. Neurosci.* **34,** 4708–4727 (2014).
80. Shouse, M. N. & Siegel, J. M. Pontine regulation of REM sleep components in cats: integrity of the pedunculopontine tegmentum (PPT) is important for phasic events but unnecessary for atonia during REM sleep. *Brain Res.* **571,** 50–63 (1992).
81. Grace, K. P., Vanstone, L. E. & Horner, R. L. Endogenous cholinergic input to the pontine REM sleep generator is not required for REM sleep to occur. *J. Neurosci.* **34,** 14198–14209 (2014).
82. Van Dort, C. J. *et al.* Optogenetic activation of cholinergic neurons in the PPT or LDT induces REM sleep. *Proc. Natl Acad. Sci. USA* **112,** 584–589 (2015).
The role of cholinergic brainstem neurons in the control of REM sleep has been intensively debated. This optogenetic study provided evidence that these neurons are involved in the induction rather than the maintenance of REM sleep.
83. Sakai, K. & Koyama, Y. Are there cholinergic and non-cholinergic paradoxical sleep-on neurones in the pons? *Neuroreport* **7,** 2449–2453 (1996).
84. Maloney, K. J., Mainville, L. & Jones, B. E. Differential c-Fos expression in cholinergic, monoaminergic, and GABAergic cell groups of the pontomesencephalic tegmentum after paradoxical sleep deprivation and recovery. *J. Neurosci.* **19,** 3057–3072 (1999).
85. Cox, J., Pinto, L. & Dan, Y. Calcium imaging of sleep-wake related neuronal activity in the dorsal pons. *Nat. Commun.* **7,** 10763 (2016).
86. Krenzer, M. *et al.* Brainstem and spinal cord circuitry regulating REM sleep and muscle atonia. *PLoS One* **6,** e24998 (2011).
87. Schenkel, E. & Siegel, J. M. REM sleep without atonia after lesions of the medial medulla. *Neurosci. Lett.* **98,** 159–165 (1989).
88. Magoun, H. W. & Rhines, R. An inhibitory mechanism in the bulbar reticular formation. *J. Neurophysiol.* **9,** 165–171 (1946).
89. Arrigoni, E. & Saper, C. B. What optogenetic stimulation is telling us (and failing to tell us) about fast neurotransmitters and neuromodulators in brain circuits for wake-sleep regulation. *Curr. Opin. Neurobiol.* **29,** 165–171 (2014).
90. Batini, C., Moruzzi, G., Palestini, M., Rossi, G. F. & Zanchetti, A. Persistent patterns of wakefulness in the pretrigeminal midpontine preparation. *Science* **128,** 30–32 (1958).
91. Magnes, J., Moruzzi, G. & Pompeiano, O. Synchronization of the EEG produced by low-frequency electrical stimulation of the region of the solitary tract. *Arch. Ital. Biol.* **99,** 33–67 (1961).
92. Eguchi, K. & Satoh, T. Characterization of the neurons in the region of solitary tract nucleus during sleep. *Physiol. Behav.* **24,** 99–102 (1980).
93. Sastre, J. P., Buda, C., Kitahama, K. & Jouvet, M. Importance of the ventrolateral region of the periaqueductal gray and adjacent tegmentum in the control of paradoxical sleep as studied by muscimol microinjections in the cat. *Neuroscience* **74,** 415–426 (1996).
94. Aston-Jones, G., Ennis, M., Pieribone, V. A., Nickell, W. T. & Shipley, M. T. The brain nucleus locus coeruleus: restricted afferent control of a broad efferent network. *Science* **234,** 734–737 (1986).
95. Clément, O. *et al.* The inhibition of the dorsal paragigantocellular reticular nucleus induces waking and the activation of all adrenergic and noradrenergic neurons: a combined pharmacological and functional neuroanatomical study. *PLoS One* **9,** e96851 (2014).
96. Borbély, A. A. A two process model of sleep regulation. *Hum. Neurobiol.* **1,** 195–204 (1982).
97. Dijk, D. J., Brunner, D. P., Beersma, D. G. & Borbély, A. A. Electroencephalogram power density and slow wave sleep as a function of prior waking and circadian phase. *Sleep* **13,** 430–440 (1990).
98. Franken, P., Chollet, D. & Tafti, M. The homeostatic regulation of sleep need is under genetic control. *J. Neurosci.* **21,** 2610–2621 (2001).
99. Ishimori, K. True cause of sleep: a hypnogenic substance as evidenced in the brain of sleep-deprived animals. *Tokyo Igakkai Zasshi* **23,** 429–457 (1909).
100. Legendre, R. & Pieron, H. Recherches sur le besoin de sommeil consécutif à une veille prolongée. *Z. Allg. Physiol.* **14,** 235–262 (1913).
101. Porkka-Heiskanen, T. *et al.* Adenosine: a mediator of the sleep-inducing effects of prolonged wakefulness. *Science* **276,** 1265–1268 (1997).

102. Porkka-Heiskanen, T., Strecker, R. E. & McCarley, R. W. Brain site-specificity of extracellular adenosine concentration changes during sleep deprivation and spontaneous sleep: an *in vivo* microdialysis study. *Neuroscience* **99,** 507–517 (2000).
103. Kalinchuk, A. V., McCarley, R. W., Stenberg, D., Porkka-Heiskanen, T. & Basheer, R. The role of cholinergic basal forebrain neurons in adenosine-mediated homeostatic control of sleep: lessons from 192 IgG-saporin lesions. *Neuroscience* **157,** 238–253 (2008).
104. Bjorness, T. E. *et al.* An adenosine-mediated glial-neuronal circuit for homeostatic sleep. *J. Neurosci.* **36,** 3709–3721 (2016).
105. Halassa, M. M. *et al.* Astrocytic modulation of sleep homeostasis and cognitive consequences of sleep loss. *Neuron* **61,** 213–219 (2009).
106. Huang, Z.-L. *et al.* Adenosine A2A, but not A1, receptors mediate the arousal effect of caffeine. *Nat. Neurosci.* **8,** 858–859 (2005).
107. Urade, Y. *et al.* Sleep regulation in adenosine A2A receptor-deficient mice. *Neurology* **61** (Suppl 6), S94–S96 (2003).
108. Kumar, S. *et al.* Adenosine A2A receptors regulate the activity of sleep regulatory GABAergic neurons in the preoptic hypothalamus. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **305,** R31–R41 (2013).
109. Scammell, T. E. *et al.* An adenosine A2a agonist increases sleep and induces Fos in ventrolateral preoptic neurons. *Neuroscience* **107,** 653–663 (2001).
110. Alam, M. A., Kumar, S., McGinty, D., Alam, M. N. & Szymusiak, R. Neuronal activity in the preoptic hypothalamus during sleep deprivation and recovery sleep. *J. Neurophysiol.* **111,** 287–299 (2014).
111. Morairty, S., Rainnie, D., McCarley, R. & Greene, R. Disinhibition of ventrolateral preoptic area sleep-active neurons by adenosine: a new mechanism for sleep promotion. *Neuroscience* **123,** 451–457 (2004).
112. Gallopin, T. *et al.* The endogenous somnogen adenosine excites a subset of sleep-promoting neurons via A2A receptors in the ventrolateral preoptic nucleus. *Neuroscience* **134,** 1377–1390 (2005).
113. Arrigoni, E., Chamberlin, N. L., Saper, C. B. & McCarley, R. W. Adenosine inhibits basal forebrain cholinergic and noncholinergic neurons *in vitro*. *Neuroscience* **140,** 403–413 (2006).
114. Liu, Z.-W. & Gao, X.-B. Adenosine inhibits activity of hypocretin/orexin neurons by the A1 receptor in the lateral hypothalamus: a possible sleep-promoting effect. *J. Neurophysiol.* **97,** 837–848 (2007).
115. Oishi, Y., Huang, Z.-L., Fredholm, B. B., Urade, Y. & Hayaishi, O. Adenosine in the tuberomammillary nucleus inhibits the histaminergic system via A1 receptors and promotes non-rapid eye movement sleep. *Proc. Natl Acad. Sci. USA* **105,** 19992–19997 (2008).
116. Stenberg, D. *et al.* Sleep and its homeostatic regulation in mice lacking the adenosine A1 receptor. *J. Sleep Res.* **12,** 283–290 (2003).
117. Bjorness, T. E., Kelly, C. L., Gao, T., Poffenberger, V. & Greene, R. W. Control and function of the homeostatic sleep response by adenosine A1 receptors. *J. Neurosci.* **29,** 1267–1276 (2009).
118. Porkka-Heiskanen, T. Sleep homeostasis. *Curr. Opin. Neurobiol.* **23,** 799–805 (2013).
119. Scammell, T. *et al.* Activation of ventrolateral preoptic neurons by the somnogen prostaglandin D2. *Proc. Natl Acad. Sci. USA* **95,** 7754–7759 (1998).
120. Mizoguchi, A. *et al.* Dominant localization of prostaglandin D receptors on arachnoid trabecular cells in mouse basal forebrain and their involvement in the regulation of non-rapid eye movement sleep. *Proc. Natl Acad. Sci. USA* **98,** 11674–11679 (2001).
121. Huber, R., Ghilardi, M. F., Massimini, M. & Tononi, G. Local sleep and learning. *Nature* **430,** 78–81 (2004).
122. Vyazovskiy, V., Borbély, A. A. & Tobler, I. Unilateral vibrissae stimulation during waking induces interhemispheric EEG asymmetry during subsequent sleep in the rat. *J. Sleep Res.* **9,** 367–371 (2000).
123. Gerashchenko, D. *et al.* Identification of a population of sleep-active cerebral cortex neurons. *Proc. Natl Acad. Sci. USA* **105,** 10227–10232 (2008). **This study reported sleep-active, nNOS expressing neurons in the cortex. As c-Fos activation correlates with sleep pressure, these neurons might be involved in the cortical expression of sleep homeostasis.**
124. Morairty, S. R. *et al.* A role for cortical nNOS/NK1 neurons in coupling homeostatic sleep drive to EEG slow wave activity. *Proc. Natl Acad. Sci. USA* **110,** 20272–20277 (2013).
125. Seress, L., Abrahám, H., Hajnal, A., Lin, H. & Totterdell, S. NOS-positive local circuit neurons are exclusively axo-dendritic cells both in the neo- and archi-cortex of the rat brain. *Brain Res.* **1056,** 183–190 (2005).
126. Benington, J. H. & Heller, H. C. REM-sleep timing is controlled homeostatically by accumulation of REM-sleep propensity in non-REM sleep. *Am. J. Physiol.* **266,** R1992–R2000 (1994).
127. Franken, P. Long-term vs. short-term processes regulating REM sleep. *J. Sleep Res.* **11,** 17–28 (2002).
128. Chou, T. C. *et al.* Critical role of dorsomedial hypothalamic nucleus in a wide range of behavioral circadian rhythms. *J. Neurosci.* **23,** 10691–10702 (2003).
129. Wurts, S. W. & Edgar, D. M. Circadian and homeostatic control of rapid eye movement (REM) sleep: promotion of REM tendency by the suprachiasmatic nucleus. *J. Neurosci.* **20,** 4300–4310 (2000).
130. Jones, J. R., Tackenberg, M. C. & McMahon, D. G. Manipulating circadian clock neuron firing rate resets molecular circadian rhythms and behavior. *Nat. Neurosci.* **18,** 373–375 (2015).
131. King, D. P. & Takahashi, J. S. Molecular genetics of circadian rhythms in mammals. *Annu. Rev. Neurosci.* **23,** 713–742 (2000).
132. Flourakis, M. *et al.* A conserved bicycle model for circadian clock control of membrane excitability. *Cell* **162,** 836–848 (2015).
133. Hattar, S., Liao, H. W., Takao, M., Berson, D. M. & Yau, K. W. Melanopsin-containing retinal ganglion cells: architecture, projections, and intrinsic photosensitivity. *Science* **295,** 1065–1070 (2002).
134. Deboer, T., Vansteensel, M. J., Détári, L. & Meijer, J. H. Sleep states alter activity of suprachiasmatic nucleus neurons. *Nat. Neurosci.* **6,** 1086–1090 (2003).
135. Bina, K. G., Rusak, B. & Semba, K. Localization of cholinergic neurons in the forebrain and brainstem that project to the suprachiasmatic nucleus of the hypothalamus in rat. *J. Comp. Neurol.* **335,** 295–307 (1993).
136. Meyer-Bernstein, E. L., Blanchard, J. H. & Morin, L. P. The serotonergic projection from the median raphe nucleus to the suprachiasmatic nucleus modulates activity phase onset, but not other circadian rhythm parameters. *Brain Res.* **755,** 112–120 (1997).
137. Aston-Jones, G., Chen, S., Zhu, Y. & Oshinsky, M. L. A neural circuit for circadian regulation of arousal. *Nat. Neurosci.* **4,** 732–738 (2001).
138. Carter, M. E. *et al.* Tuning arousal with optogenetic modulation of locus coeruleus neurons. *Nat. Neurosci.* **13,** 1526–1533 (2010).
139. Halassa, M. M. *et al.* State-dependent architecture of thalamic reticular subnetworks. *Cell* **158,** 808–821 (2014).
140. Rossi, M. *et al.* Investigation of the feeding effects of melanin concentrating hormone on food intake--action independent of galanin and the melanocortin receptors. *Brain Res.* **846,** 164–170 (1999).
141. Kemp, B., Zwinderman, A. H., Tuk, B., Kamphuisen, H. A. & Oberyé, J. J. Analysis of a sleep-dependent neuronal feedback loop: the slow-wave microcontinuity of the EEG. *IEEE Trans. Biomed. Eng.* **47,** 1185–1194 (2000).
142. Goldberger, A. L. *et al.* PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *Circulation* **101,** e215–e220 (2000).
143. Berndt, A. *et al.* Structural foundations of optogenetics: determinants of channelrhodopsin ion selectivity. *Proc. Natl Acad. Sci. USA* **113,** 822–829 (2016).
144. Buch, T. *et al.* A Cre-inducible diphtheria toxin receptor mediates cell lineage ablation after toxin administration. *Nat. Methods* **2,** 419–426 (2005).
145. Morgan, C. W., Julien, O., Unger, E. K., Shah, N. M. & Wells, J. A. Turning on caspases with genetics and small molecules. *Methods Enzymol.* **544,** 179–213 (2014).

Acknowledgements We thank Z. Zhang for providing a 24 h sleep recording from a mouse and S. Chung for helpful comments on the manuscript. This work was supported by EMBO and Human Frontier Science Program postdoctoral fellowships (to F.W.).

Author Contributions F.W. and Y.D. wrote and revised the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to Y.D. (ydan@berkeley.edu).

Reviewer Information *Nature* thanks J. Born and the other anonymous reviewer(s) for their contribution to the peer review of this work.

doi:10.1038/nature19757

Structural insight into the role of the Ton complex in energy transduction

Hervé Celia[1,2*], Nicholas Noinaj[3*], Stanislav D. Zakharov[3], Enrica Bordignon[4,5], Istvan Botos[2], Monica Santamaria[6], Travis J. Barnard[2], William A. Cramer[3], Roland Lloubes[1] & Susan K. Buchanan[2]

In Gram-negative bacteria, outer membrane transporters import nutrients by coupling to an inner membrane protein complex called the Ton complex. The Ton complex consists of TonB, ExbB, and ExbD, and uses the proton motive force at the inner membrane to transduce energy to the outer membrane via TonB. Here, we structurally characterize the Ton complex from *Escherichia coli* using X-ray crystallography, electron microscopy, double electron–electron resonance (DEER) spectroscopy, and crosslinking. Our results reveal a stoichiometry consisting of a pentamer of ExbB, a dimer of ExbD, and at least one TonB. Electrophysiology studies show that the Ton subcomplex forms pH-sensitive cation-selective channels and provide insight into the mechanism by which it may harness the proton motive force to produce energy.

Gram-negative bacteria contain no known energy source located in the outer membrane. To overcome this deficiency, bacteria have evolved systems to harness the energy produced from the proton motive force (PMF) generated at the inner membrane and to transduce it for transport at the outer membrane[1,2]. An example is the Ton system, which mediates the uptake of metals, carbohydrates, iron-siderophore complexes, cobalamin, and many bacteriocins[3–6]. The Ton system consists of two elements: an energy-transducing Ton complex located within the inner membrane, and a ligand-specific TonB-dependent transporter within the outer membrane, which interacts with the ligands to be transported[2,7–9] (Fig. 1a). The Ton complex is formed by three integral polytopic membrane proteins: ExbB, ExbD, and TonB. The energy transfer is mediated by a conserved 5–7 residue TonB-box at the N terminus of all TonB-dependent transporters[2,10]. Upon ligand binding to the TonB-dependent transporter, the TonB box becomes exposed and the interaction with TonB leads to conformational changes in the TonB-dependent transporter that are coupled to ligand transport across the outer membrane.

ExbB is predicted to contain three transmembrane spanning helices (TMHs) and a large cytoplasmic domain, whereas ExbD and TonB are each predicted to contain a single N-terminal TMH that anchors a large C-terminal periplasmic domain in the inner membrane[11–15] (Fig. 1a). The exact stoichiometry of components of the Ton complex has been a matter of debate for years[16–19]. Evidence favouring a dynamic mechanism has been reported in which fluorescence anisotropy studies showed that the presence of TonB within the Ton complex sustains a rotational motion that depends on the PMF at the inner membrane[20].

The Ton complex is often compared to the Tol complex, which consists of the analogous components TolQ, TolR, and TolA[21,22]. The Tol complex is required for cell envelope integrity[23,24] and to maintain cellular structure during cell division[25]. Similar to TonB for the Ton system, TolA has been shown to undergo energy-dependent conformational changes[20,26,27]. The Ton complex is also evolutionarily related to the Mot complex, which drives bacterial flagellar motion[22,28–30].

To better understand the role of the Ton complex in energy transduction to the outer membrane, we solved crystal structures of the *E. coli* Ton subcomplex. We further characterized the assembly of the complex using electron microscopy, crosslinking, and DEER spectroscopy, which reveal that the fully assembled Ton complex consists of a pentamer of ExbB, a dimer of ExbD, and at least one TonB.

Crystal structure of the Ton subcomplex

Constructs of the Ton subcomplex (ExbB–ExbD) were purified using a C-terminal 10×His tag on ExbD (Fig. 1b, c) and crystals grown by vapour diffusion (see Methods). Initial phases were calculated using a 5.2 Å Se-SAD (single-wavelength anomalous diffraction) dataset of ExbB–ExbD$_{\Delta peri}$, allowing an initial poly-alanine model to be built (Extended Data Fig. 1). This starting model was then used as a search model to solve the structures at pH 4.5 and 7.0 by molecular replacement (Supplementary Table 1).

The structure of the ExbB–ExbD$_{\Delta peri}$ complex at pH 7.0 was solved to 2.6 Å resolution. However, only ExbB could be built, owing to insufficient density for ExbD$_{\Delta peri}$ (Extended Data Fig. 2). The ExbB monomer adopts an extended conformation sitting perpendicular to the membrane, consisting of seven α-helices with α2 and α7 measuring 80–100 Å in length and α5 and α6 forming an extended helix (~100 Å) separated by a kink (Fig. 1d). The transmembrane domain consists of three transmembrane helices (α2, α6, and α7) which extend into the cytoplasm to form a 5-helix bundle with cytoplasmic domain 1 and the C-terminal domain.

The quaternary structure of ExbB is a pentamer in which the five transmembrane domains form a transmembrane pore (α6 and α7), while the cytoplasmic domains form a large enclosed cavity extending as far as ~60 Å into the cytoplasm (Fig. 1e–g). The cytoplasmic domain of ExbB retains five-fold symmetry with each edge measuring around 45 Å, while the periplasmic domain is arranged in pseudo-five-fold symmetry with each edge measuring around 35 Å. ExbB forms a large extended cavity (largest pore radius around 11 Å) along the cytoplasmic and transmembrane domains that is open but constricted at each end

[1]Laboratoire d'Ingénierie des Systèmes Macromoléculaires, UMR7255 CNRS/Aix-Marseille Université, Institut de Microbiologie de la Méditerranée, 13402 Marseille Cedex 20, France. [2]National Institute of Diabetes & Digestive & Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA. [3]Markey Center for Structural Biology, Department of Biological Sciences, and the Purdue Institute for Inflammation, Immunology and Infectious Diseases, Purdue University, West Lafayette, Indiana 47907, USA. [4]Fachbereich Physik, Freie Universität, 14195 Berlin, Germany. [5]Faculty of Chemistry and Biochemistry, Ruhr-Universität Bochum, 45810 Bochum, Germany. [6]Departamento de Cirugia Experimental, Instituto de Investigacion Hospital La Paz (IdiPAZ), Paseo de la Castellana 261, 28046 Madrid, Spain.
*These authors contributed equally to this work.

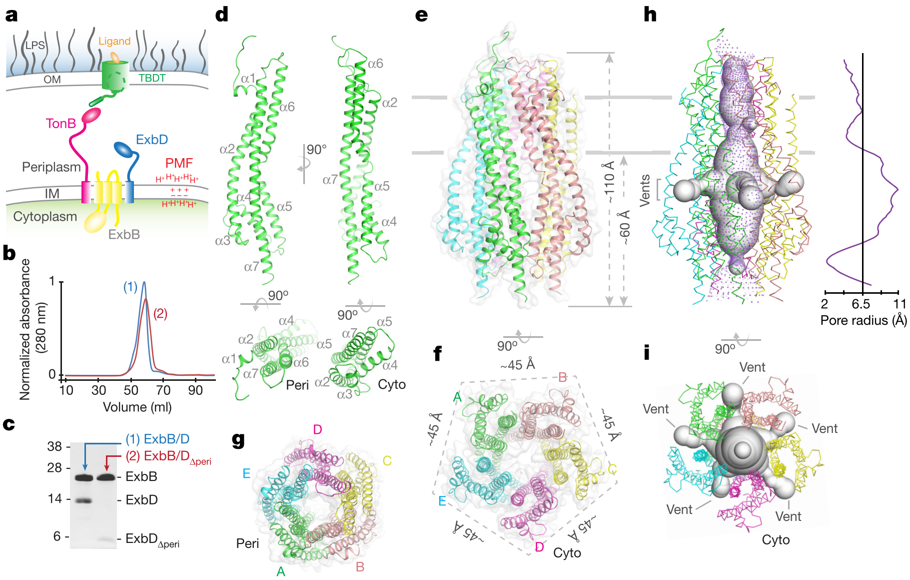


Figure 1 | The structure of the ExbB oligomer. a, Schematic of the Ton system for energy transduction. IM, inner membrane; OM, outer membrane; TBDT, TonB-dependent transporter. **b**, SEC profiles of the Ton subcomplexes (1, ExbB–ExbD; 2, ExbB–ExbD$_{\Delta peri}$; representative purification from 50 or 30 experiments, respectively). **c**, SDS–PAGE analysis of the Ton subcomplexes purified in **b**. **d**, Cartoon representation of the ExbB monomer, consisting of seven α-helices. Peri, periplasm; Cyto, cytoplasm. **e**, The ExbB pentamer structure shown as cartoon and transparent surface. **f**, Perpendicular view of the cytoplasmic end of the ExbB pentamer depicting the five-fold symmetry with each edge measuring ~45 Å. **g**, Perpendicular view of the periplasmic end of the ExbB pentamer. **h**, The ExbB pentamer was analysed with the programs MOLE 2.0[39] (spheres representation) and HOLE[40] (purple dots). **i**, Perpendicular view of the cavities shown in **h** to better illustrate the five fenestrations (vents).

(pore radius approximately 2 Å on the cytoplasmic side and 4 Å along the transmembrane side; Fig. 1h). Each monomer has approximately 3,000 $Å^2$ of buried surface area with the two adjacent molecules (about 20% of total surface area), indicating a stable oligomeric state. For the cytoplasmic cavity, five side fenestrations are observed that could allow solvent or ion passage (Fig. 1h, i). Sparse electron density indicated that the transmembrane pore of ExbB is probably filled by the TMH of ExbD$_{\Delta peri}$; however, this density was too diffuse to allow a model to be built unambiguously (Extended Data Fig. 2). Two ExbB pentamers were observed per asymmetric unit and alignment of these pentamers revealed some helical shifts, possibly indicating a propensity for movement within the membrane domain (Extended Data Fig. 3).

To verify the presence of the TMH of ExbD$_{\Delta peri}$ within the transmembrane pore of the ExbB pentamer, we solved the structure of ExbB–ExbD$_{\Delta peri}$ at pH 4.5 to 3.5 Å resolution and observed a single α-helix (Fig. 2a, b, Supplementary Table 1 and Extended Data Fig. 4). An extended α-helix could be built consisting of residues 22–45, which correlated well with the hydrophobic residues inside the transmembrane pore of ExbB, although it was offset by about 10 Å from the position of the transmembrane domains of ExbB, which are predicted to be embedded into the membrane. The exact position of each residue was less precise owing to the lack of well-defined side-chain density. These results suggest that movements of the TMH of ExbD$_{\Delta peri}$ may be modulated by changes in pH (Extended Data Fig. 5).

A striking feature of the ExbB pentamer is the very large cytoplasmic domain and its electrostatic properties, which include a strongly electropositive 'basic belt' that sits close to the membrane interface and a strongly electronegative 'cap' that sits at the cytoplasmic end of the structure (Fig. 2c–f). For the basic belt, each monomer contributes six lysine residues at positions 44, 52, 56, 81, 108, and 206 and twelve arginine residues at positions 53, 54, 57, 66, 110, 114, 117, 118, 124, 128, 200, and 222. For the cap, each monomer contributes seven aspartates at positions 73, 77, 102, 103, 211, 223, and 225, and 11 glutamate residues at positions 47, 58, 64, 90, 94, 96, 99, 105, 109, 116, and 227. Residues E105 and E109 line the cytoplasmic pore, where we observed a single calcium ion in our structure (Fig. 2c, d).

ExbB is a pentamer in the Ton complex

Negative stain electron microscopy was performed on 2D crystals of the full-length ExbB–ExbD complex (Fig. 3a). The best images were used to generate an averaged 2D projection map of the unit cell, which revealed five domains arranged as a pentamer, each with a diameter of 20–25 Å and with the edges of the pentamer measuring about 45 Å (Fig. 3b and Extended Data Fig. 6).

The Ton subcomplex was also studied using DEER spectroscopy, in which ExbB was labelled at C25 using the spin label *S*-(1-oxyl-2,2,5,5-tetramethyl-2,5-dihydro-1H-pyrrol-3-yl)methyl methanesulfonothioate (MTSL). Using this method, distance distributions were obtained experimentally and compared to simulations of the *in silico* labelled crystal structure (Fig. 3c and Extended Data Fig. 7). The experimental results agree well with the simulated distances, with peaks at approximately 35 and 50–60 Å (Fig. 3d). Together with the crystal structure and electron microscopy studies, these results further verify the stoichiometry of ExbB as a pentamer within the Ton subcomplex containing a centralized transmembrane pore (Fig. 3b, d).

To determine the oligomeric state of ExbB in the presence of TonB, the fully assembled Ton complex was expressed and purified, and found to have a larger hydrodynamic radius than the ExbB–ExbD subcomplex (Fig. 3e and Extended Data Fig. 8). We then labelled ExbB at position 25 with MTSL and repeated the DEER spectroscopy analysis. The distance distributions were nearly identical to those of the subcomplex (Fig. 3f and Extended Data Fig. 7), confirming that ExbB is a pentamer in both the absence and the presence of TonB.

ExbD is a dimer in the Ton complex

Previous studies have suggested that the Ton complex may contain a dimer of ExbD[31]. To investigate this possibility, we engineered an $ExbB_{C25S}$–$ExbD_{E113C}$ construct of the Ton subcomplex. The sample was incubated with the crosslinker 1,8-bismaleimidodiethyleneglycol (BM(PEG)$_2$) and then separated by size exclusion chromatography (SEC) and compared to a control sample that was not crosslinked (Fig. 4a). SDS–PAGE analysis confirmed the shift of ExbD from monomer to dimer for the crosslinked sample; however, no shift was induced

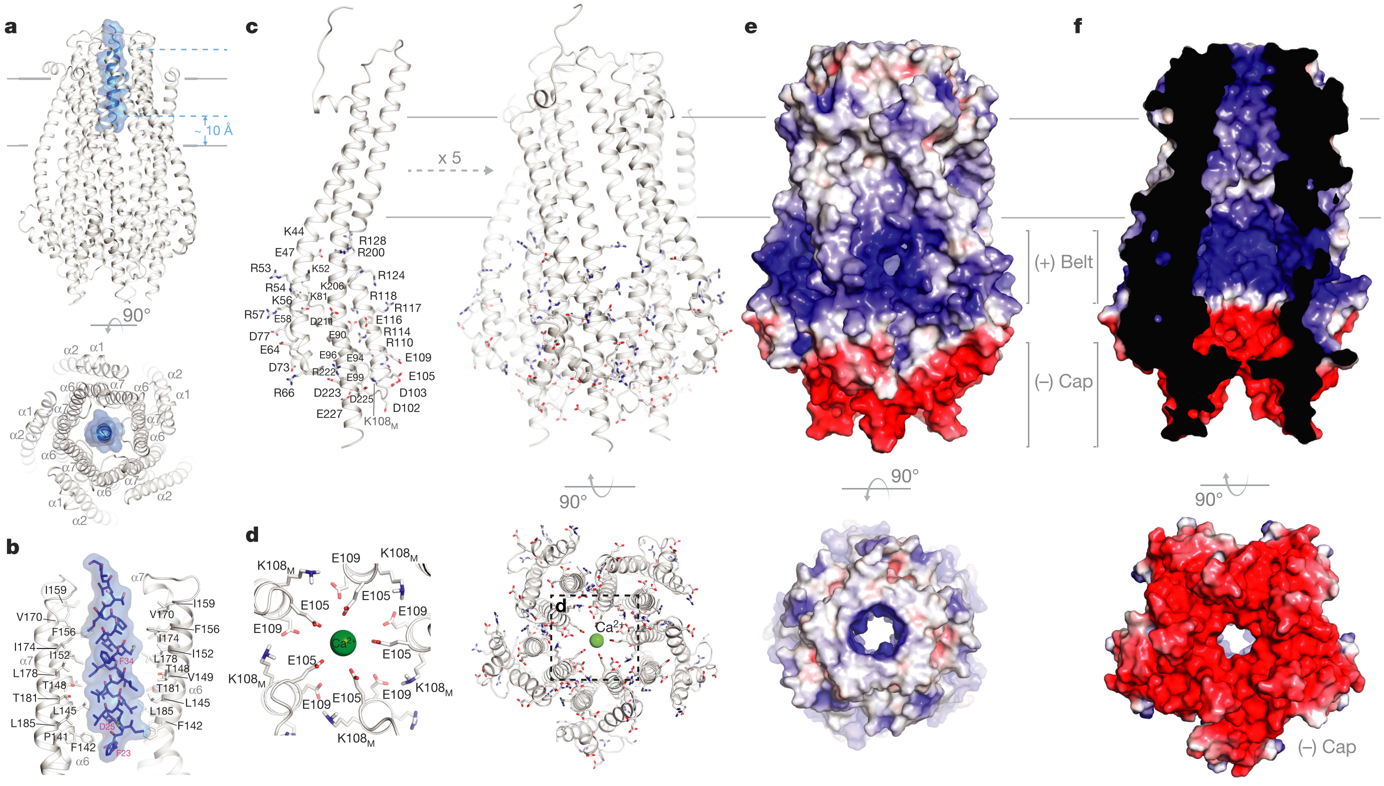


Figure 2 | The structure of the ExbB–ExbD$_{\Delta peri}$ complex. a, The ExbB–ExbD$_{\Delta peri}$ complex highlighting the transmembrane helix of ExbD (blue) located within the transmembrane pore of the ExbB pentamer (grey). **b**, Residues from helices α6 and α7 line the transmembrane pore of ExbB (grey) and mediate interactions with the transmembrane helix of ExbD (blue). For clarity, only two monomers of the ExbB pentamer are shown. **c**, The cytoplasmic domain of ExbB forms a large enclosed cavity that includes 12 arginines, 6 lysines, 11 glutamates, and 7 aspartates from each monomer (acidic residues in red, basic residues in blue). **d**, Electronegatively charged residues E105 and E109 line the cytoplasmic pore and interact with a single calcium ion (green). Left, expanded view of dashed box on right. **e**, Electrostatic surface representation of ExbB showing the electropositive 'belt' and the electronegative 'cap'. Blue and red shades indicate electropositivity (blue) or electronegativity (red). **f**, Cutaway view showing the electrostatic surface properties of the inside cavity.

by SEC, indicating that the ExbD-crosslinked dimer was formed within a single complex (intra) rather than between two different complexes (inter).

DEER spectroscopy was performed on the subcomplex by labelling ExbD at residues 78 and 113 individually, and constructs of ExbB$_{C25S}$–ExbD$_{N78C}$ and ExbB$_{C25S}$–ExbD$_{E113C}$ were labelled with MTSL. Distance distributions were detected experimentally and compared to simulations of an *in silico* labelled model of the ExbD dimer (PDB ID 2PFU)[15], which was based on the related TolR dimer structure (PDB ID 2JWK)[32] (Fig. 4b, c and Extended Data Fig. 7). According to the dimer model, labelling at residue 78 would yield distances of 32–44 Å, which is consistent with the peaks observed experimentally at 35 and 43 Å (Fig. 4b, d). Furthermore, labelling at residue 113 would yield distances of 15–35 Å, which is also consistent with the peaks observed experimentally at 23 and 34 Å, within the accuracy of the rotamer library approach (Fig. 4c, d and Extended Data Fig. 7).

To determine the oligomeric state of ExbD in the presence of TonB, DEER spectroscopy was performed on the fully assembled Ton complex containing the TonB$_{C18A}$, ExbB$_{C25S}$ and ExbD$_{N78C}$ mutations and labelled with MTSL. The distance distributions for the labels on ExbD were nearly identical to those of the subcomplex (Fig. 4e and Extended Data Fig. 7), confirming that ExbD is a dimer in both the absence and the presence of TonB.

Ton subcomplex channel properties

To investigate ion conduction by the Ton subcomplex (ExbB–ExbD)[27,33], the subcomplex was reconstituted into liposomes that were fused with a preformed planar bilayer membrane[34]. Single- and multichannel recordings revealed that channels formed by the Ton subcomplex display a conductance of 120 ± 30 pS at neutral pH (Fig. 5a, b), whereas channels formed by the ExbB pentamer are nearly twice as large with a conductance of 220 ± 50 pS (Fig. 5b). This is consistent with our structure, which shows the transmembrane helix of ExbD plugging the transmembrane pore of the ExbB pentamer.

We also determined the ion selectivity of the channels. Channels of the Ton subcomplex have a pronounced cation selectivity with sevenfold greater permeability for K^+ than for Cl^- (V_{rev}, 24.7 ± 0.9 mV; pK^+/pCl^-, 7.0 ± 0.9) (Supplementary Table 2). Channels formed by the ExbB–ExbD$_{\Delta peri}$ complex are less cation selective (V_{rev}, 13.7 ± 4.5 mV; pK^+/pCl^-, 2.6 ± 1.0), which implies that the periplasmic domain of ExbD enhances cation selectivity. However, the ExbB pentamer is anion-selective (V_{rev}, −12.6 ± 2.8 mV; pK^+/pCl^-, 0.43 ± 0.09) (Supplementary Table 2), indicating that ExbD$_{\Delta peri}$ is sufficient to serve as a cation-selective filter. The point mutation D25A in the transmembrane helix of ExbD, which sits in the pore of the ExbB pentamer, markedly decreases the cation selectivity of the Ton subcomplex (V_{rev}, 17.0 ± 1.5 mV; pK^+/pCl^-, 3.3 ± 0.5) (Supplementary Table 2), indicating that D25 makes a substantial contribution toward ion selectivity.

The channel activity of the Ton subcomplex has a pronounced pH dependence, showing a marked decrease in transmembrane current upon a decrease in pH from neutral to acidic (Fig. 5c). However, the transmembrane helix of ExbD is not the major contributor to the observed pH dependence, as the D25A mutant shows a nearly identical pH dependence to that of the wild type (Fig. 5c), suggesting that the unique electrostatic properties of the ExbB pentamer may be responsible. The decrease in transmembrane current amplitude in the pH range 4.5–8.0 is explained by a decrease in single-channel conductance from

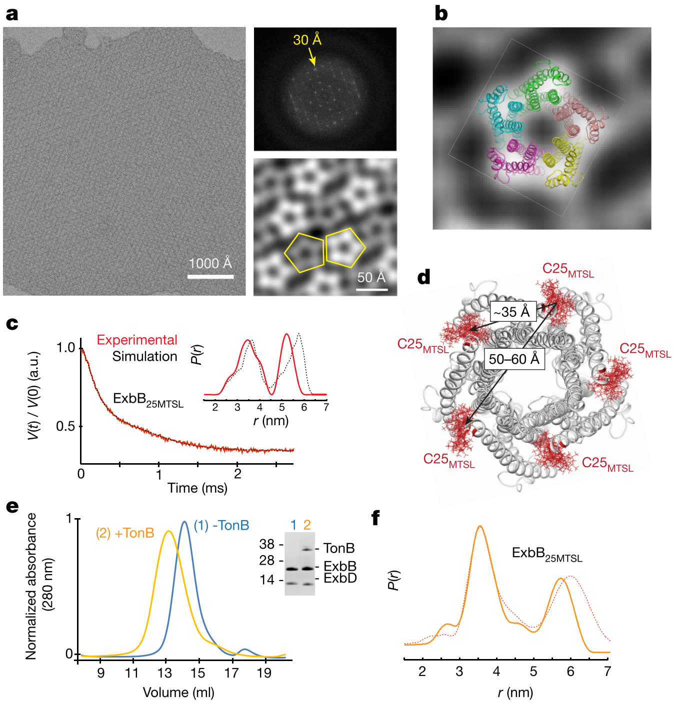


Figure 3 | The oligomeric state of ExbB within the Ton complex. **a**, Electron microscopy analysis was performed using 2D crystals (left) of the Ton subcomplex with a power spectrum out to ~30 Å (top right). Five images were analysed, and a representative averaged projection map calculated from 900 sub-images shows that the complex is pentameric (bottom right). **b**. The electron microscopy studies are consistent with ExbB being a pentamer with edges measuring ~45 Å. **c**. DEER spectroscopy was performed on the Ton subcomplex labelled with MTSL at position C25 of ExbB. The experimentally measured traces and distance distributions (inset, red lines) agree well with those calculated from the *in silico* labelled ExbB (black dashed lines). **d**, DEER analyses of the Ton subcomplex are consistent with ExbB being a pentamer. **e**, Purification of the fully assembled Ton complex (orange) compared with the subcomplex (blue). **f**, Comparison of distance distributions of the fully assembled Ton complex (solid orange line) to those of the Ton subcomplex in DDM lacking TonB (dashed red line) showed minimal differences. **c**, **e**, and **f** show data from single experiments.

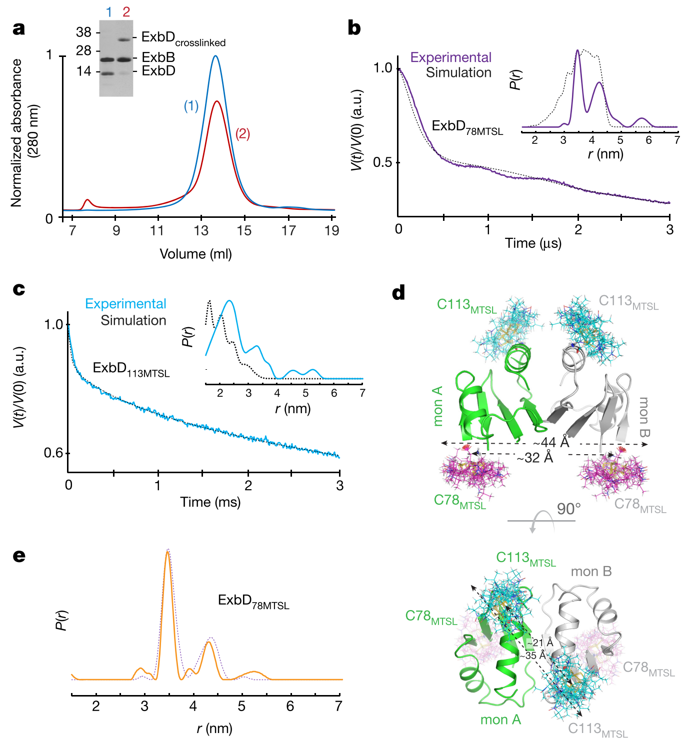


Figure 4 | The oligomeric state of ExbD within the Ton complex. **a**, Crosslinking studies targeting ExbD are consistent with a dimer within the Ton subcomplex, as evidenced by an observed crosslinked dimer (red, lane 2) compared to the non-crosslinked sample (blue, lane 1). **b**, **c**, DEER spectroscopy was performed on ExbD labelled at position 78 (purple lines, **b**) and position 113 (cyan lines, **c**). The experimentally measured traces and distance distributions (insets, purple and cyan lines) are consistent with those calculated (black dashed lines) from the *in silico* labelled ExbD dimer model (PDB ID 2PFU), which is based on the reported TolR structure (PDB ID 2JWK). **d**, The distance measurements within the *in silico* labelled ExbD dimer model are in agreement with those obtained experimentally at each site using DEER analysis. **e**, DEER spectroscopy was performed in DDM on the fully assembled Ton complex labelled at position 78 on ExbD. Comparison of distance distributions of the fully assembled Ton complex (solid orange line) to those of the Ton subcomplex in DDM lacking TonB (dashed purple line) show minimal differences. **a**–**c** and **e** show data from single experiments.

120 pS at pH 8.0 to 70 pS at pH 4.5 (Fig. 5d). Below pH 4.5, the decrease in transmembrane current is caused by channel closure at both positive and negative potentials (Fig. 5c). The ion channel conductance properties of the Ton subcomplex demonstrate that it is being modulated by pH, possibly through movement of the transmembrane helix of ExbD within the transmembrane pore of the ExbB pentamer, such that at low pH, the transmembrane helix of ExbD is in a more closed/fixed conformation (Fig. 2b).

Model of a fully assembled Ton complex

On the basis of our findings, we propose a model in which the Ton complex consists of a pentamer of ExbB, a dimer of ExbD, and at least one TonB (Fig. 5e and Extended Data Fig. 9). As only a single transmembrane helix can fit within the transmembrane pore of ExbB, and dimerization of ExbD is hypothesized to be mediated by its periplasmic domain, we propose that a second copy of ExbD is located outside the ExbB pentamer. Previous studies have indicated that TonB may exchange for one of the ExbD monomers during energy transduction[17]. However, our studies show that association of TonB does not notably affect the structure or stoichiometry of ExbB or ExbD within the Ton complex. The interaction of TonB with ExbD leads to a functional Ton complex, triggering energy production and transduction in the form of conformational changes in TonB that lead to ligand uptake by the transporter at the outer membrane[35,36].

The Ton complex relies on the PMF for its function[27,33] and it has been proposed that the Ton complex acts as a proton-conducting channel that shuttles protons from the periplasm to the cytoplasm and that this powers a mechanical motion within the complex[20]. Mutagenesis studies have previously identified a number of residues that are necessary for harnessing the PMF, including D25 of ExbD and T148 and T181 of ExbB[37,38]. These residues all map to the interior of the transmembrane pore of ExbB, where protons would be translocated (Extended Data Fig. 10). Our studies indicate that the transmembrane helix of ExbD is quite dynamic within the transmembrane pore of ExbB, and together with the electrophysiology experiments, show that this dynamic behaviour can be modulated by pH. The electrostatics of the ExbB pentamer may also create an 'electrostatic funneling' effect that helps to draw protons from the periplasm and steer them through the transmembrane pore of ExbB into the cytoplasm (Fig. 5f). Therefore, we suggest two plausible mechanistic models for how the Ton complex harnesses the PMF for energy production and transduction (Fig. 5g). The first is the 'electrostatic piston' model, in which the transmembrane helix of ExbD moves translationally within the transmembrane pore of ExbB, thereby creating a piston-like motion. The second is the 'rotational' model, in which the transmembrane helix of ExbD rotates within the transmembrane pore of ExbB, creating rotational motion.

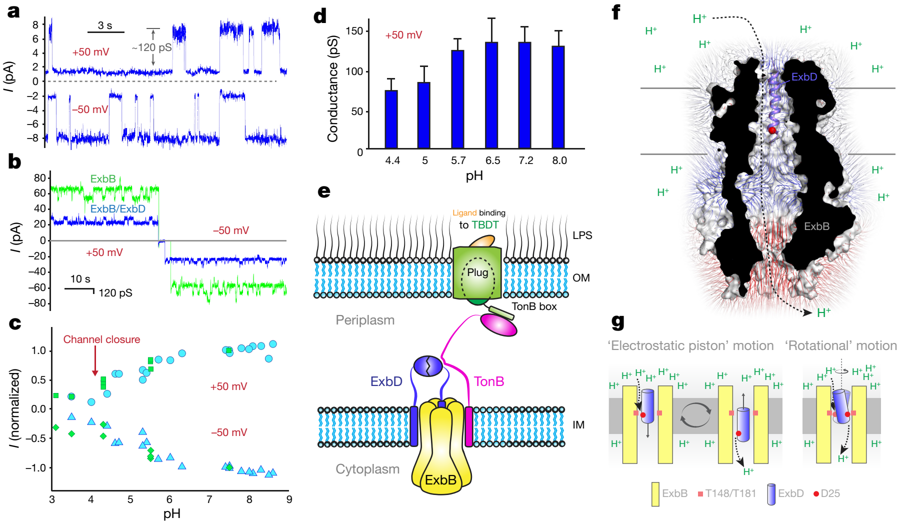


Figure 5 | Channel properties of the Ton subcomplex. a, Representative spectra for single-channel measurements of the Ton subcomplex ($n = 15$). **b**, Representative spectra of multichannel measurements performed on the Ton subcomplex (blue) and ExbB alone (green; $n = 15$ for each sample). **c**, Dependence of the macroscopic current amplitude on pH for the Ton subcomplex (blue) and the D25A mutation in the TM helix of ExbD (green) with a holding potential of +50 mV (circles and squares) or −50 mV (triangles and diamonds). **d**, Dependence of single-channel conductance of the Ton subcomplex on pH measured at +50 mV. Bars show mean ± s.d. with $n = 5$. **e**, Structural model of the Ton complex consisting of a pentamer of ExbB, a dimer of ExbD, and at least one TonB. LPS, lipopolysaccharide. **f**, The ExbB–ExbD$_{\Delta\text{peri}}$ complex showing electric fields for the ExbB pentamer. The modelled location of residue D25 of the transmembrane helix of ExbD (purple) is indicated as a red sphere. **g**, Proposed mechanisms for how the Ton subcomplex may use the PMF to produce energy.

A combination of the two mechanistic models is also plausible. While we observe minor conformational shifts within the transmembrane helices of ExbB in our structures, it is also feasible that the ExbB pentamer cycles through more pronounced conformations to either drive or accommodate the dynamics of the transmembrane helix of ExbD.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 29 December 2015; accepted 15 August 2016.
Published online 21 September; corrected online 5 October 2016 (see full-text HTML version for details).

1. Krewulak, K. D. & Vogel, H. J. TonB or not TonB: is that the question? *Biochem. Cell Biol.* **89,** 87–97 (2011).
2. Noinaj, N., Guillier, M., Barnard, T. J. & Buchanan, S. K. TonB-dependent transporters: regulation, structure, and function. *Annu. Rev. Microbiol.* **64,** 43–60 (2010).
3. Bassford, P. J., Jr, Bradbeer, C., Kadner, R. J. & Schnaitman, C. A. Transport of vitamin B12 in tonB mutants of *Escherichia coli*. J. Bacteriol. **128,** 242–247 (1976).
4. Lohmiller, S., Hantke, K., Patzer, S. I. & Braun, V. TonB-dependent maltose transport by *Caulobacter crescentus*. *Microbiology* **154,** 1748–1754 (2008).
5. Schauer, K., Rodionov, D. A. & de Reuse, H. New substrates for TonB-dependent transport: do we only see the 'tip of the iceberg'? *Trends Biochem. Sci.* **33,** 330–338 (2008).
6. Cascales, E. *et al.* Colicin biology. *Microbiol. Mol. Biol. Rev.* **71,** 158–229 (2007).
7. Postle, K. & Kadner, R. J. Touch and go: tying TonB to transport.*Mol. Microbiol.* **49,** 869–882 (2003).
8. Clément, E., Mesini, P. J., Pattus, F. & Schalk, I. J. The binding mechanism of pyoverdin with the outer membrane receptor FpvA in *Pseudomonas aeruginosa* is dependent on its iron-loaded status. *Biochemistry* **43,** 7954–7965 (2004).
9. Cadieux, N., Barekzi, N. & Bradbeer, C. Observations on the calcium dependence and reversibility of cobalamin transport across the outer membrane of *Escherichia coli*. *J. Biol. Chem.* **282,** 34921–34928 (2007).
10. Schramm, E., Mende, J., Braun, V. & Kamp, R. M. Nucleotide sequence of the colicin B activity gene *cba*: consensus pentapeptide among TonB-dependent colicins and receptors. *J. Bacteriol.* **169,** 3350–3357 (1987).
11. Brewer, S. *et al.* Structure and function of X-Pro dipeptide repeats in the TonB proteins of *Salmonella typhimurium* and *Escherichia coli*. *J. Mol. Biol.* **216,** 883–895 (1990).
12. Chang, C., Mooser, A., Plückthun, A. & Wlodawer, A. Crystal structure of the dimeric C-terminal domain of TonB reveals a novel fold. *J. Biol. Chem.* **276,** 27535–27540 (2001).
13. Witty, M. *et al.* Structure of the periplasmic domain of *Pseudomonas aeruginosa* TolA: evidence for an evolutionary relationship with the TonB transporter protein. *EMBO J.* **21,** 4207–4218 (2002).
14.Ködding, J. *et al.* Crystal structure of a 92-residue C-terminal fragment of TonB from *Escherichia coli* reveals significant conformational changes compared to structures of smaller TonB fragments. *J. Biol. Chem.* **280,** 3022–3028 (2005).
15. Garcia-Herrero, A., Peacock, R. S., Howard, S. P. & Vogel, H. J. The solution structure of the periplasmic domain of the TonB system ExbD protein reveals an unexpected structural homology with siderophore-binding proteins. *Mol. Microbiol.* **66,** 872–889 (2007).
16. Higgs, P. I., Larsen, R. A. & Postle, K. Quantification of known components of the *Escherichia coli* TonB energy transduction system: TonB, ExbB, ExbD and FepA. *Mol. Microbiol.* **44,** 271–281 (2002).
17. Sverzhinsky, A. *et al.* Membrane protein complex ExbB4-ExbD1-TonB1 from *Escherichia coli* demonstrates conformational plasticity. *J. Bacteriol.* **197,** 1873–1885 (2015).
18. Sverzhinsky, A. *et al.* Coordinated rearrangements between cytoplasmic and periplasmic domains of the membrane protein complex ExbB-ExbD of *Escherichia coli*. *Structure* **22,** 791–797 (2014).
19. Pramanik, A. *et al.* Oligomeric structure of ExbB and ExbB-ExbD isolated from *Escherichia coli* as revealed by LILBID mass spectrometry. *Biochemistry* **50,** 8950–8956 (2011).
20. Jordan, L. D. *et al.* Energy-dependent motion of TonB in the Gram-negative bacterial inner membrane. *Proc. Natl Acad. Sci. USA* **110,** 11553–11558 (2013).
21. Braun, V. The structurally related exbB and tolQ genes are interchangeable in conferring tonB-dependent colicin, bacteriophage, and albomycin sensitivity. *J. Bacteriol.* **171,** 6387–6390 (1989).
22. Cascales, E., Lloubès, R. & Sturgis, J. N. The TolQ-TolR proteins energize TolA and share homologies with the flagellar motor proteins MotA-MotB.*Mol. Microbiol.* **42,** 795–807 (2001).
23. Lloubès, R. *et al.* The Tol-Pal proteins of the *Escherichia coli* cell envelope: an energized system required for outer membrane integrity? *Res. Microbiol.* **152,** 523–529 (2001).
24. Yeh, Y. C., Comolli, L. R., Downing, K. H., Shapiro, L. & McAdams, H. H. The caulobacter Tol-Pal complex is essential for outer membrane integrity and the positioning of a polar localization factor. *J. Bacteriol.* **192,** 4847–4858 (2010).
25. Gray, A. N. *et al.* Coordination of peptidoglycan synthesis and outer membrane constriction during *Escherichia coli* cell division. *eLife* **4,** 07118 (2015).
26. Germon, P., Ray, M. C., Vianney, A. & Lazzaroni, J. C. Energy-dependent conformational change in the TolA protein of *Escherichia coli* involves its N-terminal domain, TolQ, and TolR. *J. Bacteriol.* **183,** 4110–4114 (2001).
27. Larsen, R. A., Thomas, M. G. & Postle, K. Protonmotive force, ExbB and ligand-bound FepA drive conformational changes in TonB. *Mol. Microbiol.* **31,** 1809–1824 (1999).
28. Zhai, Y. F., Heijne, W. & Saier, M. H., Jr. Molecular modeling of the bacterial outer membrane receptor energizer, ExbBD/TonB, based on homology with the flagellar motor, MotAB. *Biochim. Biophys. Acta* **1614,** 201–210 (2003).
29. Kojima, S. & Blair, D. F. Conformational change in the stator of the bacterial flagellar motor. *Biochemistry* **40,** 13041–13050 (2001).
30. Saier, M. H., Jr. Tracing pathways of transport protein evolution.*Mol. Microbiol.* **48,** 1145–1156 (2003).

31. Ollis, A. A., Manning, M., Held, K. G. & Postle, K. Cytoplasmic membrane protonmotive force energizes periplasmic interactions between ExbD and TonB. *Mol. Microbiol.* **73,** 466–481 (2009).
32. Parsons, L. M., Grishaev, A. & Bax, A. The periplasmic domain of TolR from *Haemophilus influenzae* forms a dimer with a large hydrophobic groove: NMR solution structure and comparison to SAXS data. *Biochemistry* **47,** 3131–3142 (2008).
33. Larsen, R. A. & Postle, K. Conserved residues Ser(16) and His(20) and their relative positioning are essential for TonB activity, cross-linking of TonB with ExbB, and the ability of TonB to respond to proton motive force. *J. Biol. Chem.* **276,** 8111–8117 (2001).
34. Labarca, P. & Latorre, R. Insertion of ion channels into planar lipid bilayers by vesicle fusion. *Methods Enzymol*. **207,** 447–463 (1992).
35. Ollis, A. A., Kumar, A. & Postle, K. The ExbD periplasmic domain contains distinct functional regions for two stages in TonB energization. *J. Bacteriol.* **194,** 3069–3077 (2012).
36. Ollis, A. A. & Postle, K. ExbD mutants define initial stages in TonB energization. *J. Mol. Biol.* **415,** 237–247 (2012).
37. Braun, V. *et al.* Energy-coupled transport across the outer membrane of *Escherichia coli*: ExbB binds ExbD and TonB *in vitro*, and leucine 132 in the periplasmic region and aspartate 25 in the transmembrane region are important for ExbD activity. *J. Bacteriol*. **178,** 2836–2845 (1996).
38. Braun, V. & Herrmann, C. Point mutations in transmembrane helices 2 and 3 of ExbB and TolQ affect their activities in *Escherichia coli* K-12. *J. Bacteriol*. **186,** 4402–4406 (2004).
39. Sehnal, D. *et al.* MOLE 2.0: advanced approach for analysis of biomacromolecular channels. *J. Cheminform*. **5,** 39 (2013).
40. Smart, O. S., Neduvelil, J. G., Wang, X., Wallace, B. A. & Sansom, M. S. HOLE: a program for the analysis of the pore dimensions of ion channel structural models. *J. Mol. Graph.* **14,** 354–360, 376 (1996).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank the staff at the SER-CAT and GM/CA-CAT beamlines at the Advanced Photon Source (APS), Argonne National Laboratory (use of the APS is supported by the US DOE, Office of Science, Office of Basic Energy Sciences, under Contract No. W-31-109-Eng-38, and by the US DOE, Basic Energy Sciences, Office of Science, under contract No. DE-AC02-06CH11357); the staff at beamlines 5.0.1 and 8.2.1, Advance Light Source at Lawrence Berkeley National Laboratory for their assistance during crystal screening (the Advanced Light Source is supported by the Director, Office of Science, Office of Basic Energy Sciences, of the US DOE under Contract No. DE-AC02-05CH11231); G. Jeschke (ETH Zurich) for providing the Q-band resonator; T. Assafa for the reproducibility measurements; the Deutsche Forschungsgemeinschaft for funding the AWG E580 Q-band spectrometer (INST 130/972-1 FUGG); Y. Li at the NINDS/NIH Protein/Peptide Sequencing Facility for performing mass spectrometry analysis; and members of the Lloubes team, E. Cascales, J. Sturgis and J. P. Duneau for discussions. N.N. is supported by the Department of Biological Sciences, Purdue University and by the National Institute of Allergy and Infectious Diseases (1K22AI113078-01). E.B. is supported by the Cluster of Excellence RESOLV (EXC 1069) funded by the Deutsche Forschungsgemeinschaft. W.A.C. is supported by the NIH (NIH GM 038323) and the Henry Koffler Professorship at Purdue University. H.C. and R.L. are funded by the Centre National de la Recherche Scientifique, the Aix-Marseille Université and grants from the Agence National de la Recherche (BACMOLMOT [ANR-14-CE09-0023]) and from Projets internationaux de coopération scientifique (PICS05853). H.C., T.J.B., I.B. and S.K.B. are supported by the Intramural Research Program of the NIH, NIDDK.

Author Contributions M.S. prepared the ExbB wild-type construct. H.C., N.N., T.J.B., and R.L. prepared ExbD constructs and mutants of ExbB and ExbD. T.J.B. prepared the TonB constructs. H.C. performed all expression and purification for assays, electron microscopy, DEER spectroscopy, electrophysiology, and crystallization. H.C. and N.N. performed crystallization and H.C., N.N., and I.B. did screening and data collection. N.N. solved the initial crystal structure using crystals grown by H.C. H.C. performed the electron microscopy experiments. E.B. performed the DEER spectroscopy experiments. S.D.Z. and W.A.C. performed the electrophysiology experiments. All authors analysed and discussed the data. R.L., W.A.C., and S.K.B. conceived the original projects. H.C., N.N., S.D.Z., W.A.C., R.L. and S.K.B. contributed to writing the manuscript.

Author Information Coordinates and structure factors for the ExbB/ExbD complexes have been deposited into the Protein Data Bank (PDB accession codes 5SV0 and 5SV1). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to N.N. (nnoinaj@purdue.edu), R.L. (lloubes@imm.cnrs.fr) or S.K.B. (skbuchan@helix.nih.gov).

Reviewer Information *Nature* thanks H. Mchaourab, T. Walz and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Cloning of *E. coli* (K-12 strain) ExbB, ExbD, and TonB constructs and mutants. The ExbB construct with and without a C-terminal 6×His tag was subcloned into pET26b (Novagen). ExbD was subcloned into pACYCDuet-1 vector (Novagen) with an N-terminal Strep-tag and a C-terminal 10×His tag. ExbD was also subcloned into a pCDF-1b vector (Novagen) containing a C-terminal TEV protease site followed by a 10×His tag. An ExbD$_{\Delta peri}$ construct containing a C-terminal TEV protease site (preceded by a Gly-Gly-Gly linker for efficient digestion by TEV protease) followed by a 10×His tag was constructed by deletion of the sequence encoding the periplasmic domain of ExbD (residues 50–141). TonB was cloned into a pACYCDUET-1 vector with an N-terminal 10×His tag followed by a TEV protease site. Mutants of TonB (C18A), ExbD (D25A, N78C and E113C), and ExbB (C25S) were prepared by site-directed mutagenesis (primer sequences for all cloning and mutagenesis experiments are available upon request). The sequences of all plasmid constructs and mutations were verified by sequence analysis (Macrogen USA and Eurofins Genomics GmbH).

Expression and purification of the Ton complex, subcomplexes, and components. Expression of ExbB with a C-terminal 6×His tag was performed by transforming *E. coli* BL21(DE3) cells (NEB) with the pET26b/ExbB vector. Co-expression was performed by co-transforming *E. coli* BL21(DE3) cells with the respective ExbB, ExbD, and/or TonB plasmids. For all transformations, cells were plated onto LB agar plates supplemented with appropriate antibiotics. Colonies were then used for a starter culture to inoculate 12 flasks containing either 1 l 2×YT medium (Ton subcomplex) or SelenoMet medium supplemented with L-methionine at 40 mg/l (Molecular Dimensions) (Ton complex), with appropriate antibiotics. Cultures were grown at 37 °C with shaking at 220 r.p.m. until they reached an OD_{600} of 0.5–1.0, induced with isopropyl β-D-1-thiogalactopyranoside (IPTG) to 0.1 mM final concentration, and then allowed to continue to grow overnight at 28 °C. For selenomethionine-substituted samples for experimental phasing, B834(DE3) cells (NEB) were co-transformed with pET26b/ExbB$_{C25S}$ and pCDF-1b/ExbD$_{\Delta peri}$ plasmids. Single colonies were used to inoculate 12 flasks containing 1 l SelenoMet medium (Molecular Dimensions) supplemented with 40 mg/ml L-selenomethionine and appropriate antibiotics. Cultures were grown at 37 °C with shaking at 220 r.p.m. until they reached an OD_{600} of 0.5–1.0, induced with IPTG to 0.1 mM final concentration, and then allowed to continue to grow overnight at 28 °C. Cells were harvested and used immediately or stored at −80 °C.

For purification, cells were resuspended in either 1×PBS (Ton subcomplex) or TBS (Ton complex) supplemented with 100 μM 4-(2-aminoethyl)benzenesulfonyl fluoride (AEBSF), 100 μM DNase, and 50 μg/ml lysozyme, and disrupted with two passages through an EmulsiFlex-C3 (Avestin) operating at ~15,000 p.s.i. Membranes were pelleted by ultracentrifugation in a Type 45 Ti Beckman rotor at 200,000*g* for 1 h at 4 °C. Membranes were then resuspended in 1×PBS or TBS using a dounce homogenizer and solubilized by the addition of Triton X-100 (Ton subcomplex) or DDM (Anatrace) (Ton complex) to a final concentration of 1% by stirring at medium speed for 1 h to overnight at 4 °C. Insoluble material was pelleted by ultracentrifugation in a Type 45 Ti Beckman rotor at 200,000*g* for 1 h at 4 °C and the supernatant was used immediately.

Immobilized metal affinity chromatography (IMAC) was performed on an AkTA Purifier (GE Healthcare) using a 15-ml Ni-NTA agarose column (Qiagen) equilibrated with 1×PBS or TBS supplemented with 0.1% Triton X-100 or 0.1% DDM. The supernatant was supplemented with 10 mM imidazole and loaded onto the column. The column was washed in three steps with 1×PBS or TBS supplemented with 20, 40 and 60 mM imidazole, respectively, and eluted with 1×PBS or TBS supplemented with 250 mM imidazole in 2-ml fractions. Fractions were analysed by SDS–PAGE and those fractions containing the complex were pooled. To remove the 10×His tag, TEV protease was added to the sample at 0.1 mg/ml final concentration and rocked overnight at 4 °C. For the Ton complex, the sample was then diluted 2–3 times with 25 mM HEPES, pH 7.3, and 0.1% DDM and loaded onto an anion exchange 6-ml ResourceQ column (GE Healthcare). Elution was performed with a 0–1 M NaCl gradient over 5 column volumes. For the Ton subcomplex, the sample was concentrated using an Amicon Ultra-15 Centrifugal Filter Unit with a 50-kDa MW cut-off (Millipore), filtered, and purified by size-exclusion chromatography using a Superdex 200 HL 16/600 column (GE Healthcare) at a flow rate of 0.5–1.0 ml/min. The buffer consisted of 20 mM HEPES-NaOH, pH 7.0, 150 mM NaCl, 0.01% NaN_3, and 0.08% $C_{10}E_5$. For the Ton complex, eluted fractions were concentrated using an Amicon Ultra-15 Centrifugal Filter Unit with a 100-kDa MW cut-off (Millipore), and passed over a Superose6HR 10/30 column (GE Healthcare) at a flow rate of 0.5 ml/min using 20 mM HEPES-NaOH, pH 7.0, 150 mM NaCl, and 0.05% DDM.

Densitometry analysis was performed using ImageJ software[41].

Circular dichroism. Far-UV circular dichroism (CD) spectra (185–260 nm) were measured in 0.1 M NaP_i, pH 7.0, and 0.03% DDM using quartz cuvettes with a 0.02–0.2 mm optical path length. The results were analysed using the DichroWeb package of programs[42] and different sets of reference proteins, including the SMP180 set of membrane proteins. The analysis of the thermal stability of the complexes reconstituted into liposomes was measured by the temperature dependence of the CD signal amplitude at 222 nm. Thermal melting was performed in a magnetically stirred 1-cm quartz cuvette containing 10 mM HEPES, pH 7.0, and 100 mM NaCl with a rate of temperature increase of 0.5 °C/min. Melting curves were normalized to the measured value of the molar ellipticity change at 10 °C.

Crystallization and data collection. For crystallization, samples were concentrated to ~10 mg/ml and sparse matrix screening was performed using a TTP Labtech Mosquito crystallization robot using hanging drop vapour diffusion and plates incubated at 15–21 °C. Initially, many lead conditions were observed to produce crystals with hexagonal morphology; however, none diffracted to better than ~7 Å and most suffered from anisotropy. To avoid this packing, we performed reductive methylation of our samples before crystallization using the Reductive Alkylation Kit (Hampton Research), followed by an additional size-exclusion chromatography step. This led to a condition which produced diffraction spots to ~4 Å resolution. Further optimization and screening allowed us to grow crystals in 100 mM Na-acetate, pH 4.5, 100 mM $MgCl_2$, and 25% PEG 400 that routinely diffracted to ~3.5 Å resolution or better. For heavy atom soaking, crystals were transferred to a drop containing 1 mM $HgCl_2$ and incubated overnight at room temperature and then harvested directly from the soaking condition. The best native crystals for the ExbB–ExbD$_{\Delta peri}$ complex, however, were grown from 100 mM HEPES-NaOH, pH 7.0, 100 mM $CaCl_2$, and 22% PEG MME 550 and diffracted to 2.6 Å resolution; these crystals were also used for heavy atom soaking experiments. Unfortunately, none of the heavy atom soaked crystals (nor the selenomethionine substituted crystals) were useful for phasing owing to crystal pathologies, which we suspected were twinning related. However, selenomethionine substituted crystals of the ExbB$_{C25S}$–ExbD$_{\Delta peri}$ complex were obtained using 100 mM MES/imidazole, pH 6.5, 30 mM $MgCl_2$, 30 mM $CaCl_2$, 50% ethylene glycol, and 8% PEG 8000 and diffracted to 5.2 Å resolution with no twinning-related issues. Both native and selenomethionine-substituted crystals were harvested directly from the crystallization drops. Screening for diffraction quality was performed at the GM/CA-CAT and SER-CAT beamlines at the Advanced Photon Source at Argonne National Laboratory and at beamlines 5.0.1 and 8.2.1 at the Advanced Light Source at Lawrence Berkeley National Laboratory. Final datasets were collected at the SER-CAT beamline and all data were processed using either HKL2000[43] or Xia2[44]. A summary of the data collection statistics can be found in Supplementary Table 1. The presence of both components of the Ton subcomplex within the crystals was confirmed by SDS–PAGE and mass spectrometry analyses of harvested crystals.

Structure determination. For phasing the ExbB–ExbD$_{\Delta peri}$ complex structure, three datasets were collected on selenomethionine substituted crystals of the ExbB$_{C25S}$–ExbD$_{\Delta peri}$ complex at a wavelength of 0.979 Å. The data were processed with Xia2[44] and, based on non-isomorphism, one dataset was removed. The final two datasets were processed together in space group $P4_32_12$ to a final resolution of 5.2 Å. Selenium sites (35 total) were located using HKL2MAP[45] after 5,000 tries within SHELXD at a resolution range of 20–6 Å. The sites were then fed into AutoSol (PHENIX)[46] which removed one site, producing a phase-extended density-modified electron density map into which we could build an initial poly-alanine model. Five-fold symmetry was clearly observed, with each monomer consisting of very elongated α-helices, and directionality was determined on the basis of the predicted topology of ExbB, which contains a single large cytoplasmic domain. This model was then used as a search model to solve the native and Hg-soaked structures by molecular replacement using PHASER/PHENIX[46,47] and the sequence docked on the basis of anomalous peaks from the SeSAD dataset. The ExbB–ExbD$_{\Delta peri}$ complex was solved in space group $P2_1$ to 2.6 Å resolution with R/R_{free} values of 0.21/0.26 and the Hg-soaked structure in space group $P2_12_12_1$ to 3.5 Å resolution with R/R_{free} values of 0.25/0.30. All model building was performed using COOT and subsequent refinement done in PHENIX[46]. r.m.s.d. analysis was performed within PyMOL (Schrödinger). Electrostatic surface properties (calculated using the Linearized Poisson-Boltzman Equation mode with a solvent radius of 1.4), including generation of the electric field lines, were analysed and visualized using the APBS plugin within PyMOL (Schrödinger). Buried surface area was calculated using the PDBePISA server[48]. Structure-related figures were made with PyMOL (Schrödinger) and Chimera[49] and annotated and finalized with Adobe Photoshop and Illustrator.

Data availability. Coordinates and structure factors for the ExbB/ExbD complexes have been deposited into the Protein Data Bank (PDB accession codes 5SV0 and 5SV1).

2D crystallization. For 2D crystallization experiments, the Ton subcomplex (ExbB–ExbD) was extracted and purified by IMAC as previously described. The sample was passed over a Superose 12 HR 10/30 column using 20 mM Tris-HCl,

pH 7, 150 mM NaCl, 0.01% NaN_3, and 0.035% Triton X-100. The purified complex was then mixed with a solution stock of *E. coli* polar lipid (Avanti Polar Lipids, Inc.) at 10 mg/ml in 2% Triton X-100, to reach final concentrations of 0.5–1.0 mg/ml protein and 0.1–0.4 mg/ml lipid. The lipid-protein-detergent samples solutions were placed into Mini Slide-A-Lyser dialysis devices (Pierce) with a 20-kDa MW cutoff, and dialysed in 1 l of 25 mM Tris-HCl, pH 7.0, 150 mM NaCl, and 0.01% NaN_3 at 4 °C. Aliquots of dialysed samples were observed periodically by electron microscopy to monitor the formation of 2D crystals.

Electron microscopy. Sample preparation for electron microscopy was carried out by applying a 5-μl drop of protein-lipid material on a glow discharged carbon-coated electron microscopy grid. Staining was performed by addition of 1% (w/v) uranyl acetate and incubation for 1 min. Grids were then imaged on a Tecnai G2 200 LaB6 electron microscope operating at 200 kV at the Institut de Microbiologie de la Méditerranée. Images were recorded with a 2K Eagle CCD camera.

The best 2D crystals were selected through observation of the power spectrum of the images using ImageJ software[41]. Selected images were processed using the IPLT Correlation Averaging suite program[50]. A filtered image was generated by optical filtering of the low resolution spots, and padded to contain only 4–6 unit cells. The padded image was cross-correlated with the original large image. The positions of the cross-correlation peaks were determined and used to extract sub-images that were summed to generate an average image of the 2D unit cell.

DEER spectroscopy. Site-directed spin labelling was used to covalently attach the spin label (1-oxyl-2,2,5,5-tetramethyl-Δ3-pyrroline-3-methyl) methanethiosulfonate (MTSL) (Toronto Research Chemicals) to Cys25 on ExbB and to cysteines engineered at positions 78 and 113 on ExbD (N78C, E113C; ExbD constructs were in the pACYC vector containing an N-terminal strep-tag and a C-terminal 10×His tag for the Ton subcomplex, and in the pCDF-1b vector for the Ton complex). For labelling with MTSL, samples were first incubated with 2–10 mM dithiothreitol (DTT) for 1–2 h and the DTT then removed by passage over a HiTrap desalting column (GE Healthcare) or during anion exchange (Ton complex). Samples were then incubated with a 10× molar excess of MTSL overnight at 4 °C and then passed over a Superose 6HR 10/30 gel filtration column (GE Healthcare) using 20 mM HEPES-NaOH, pH 7.5, 200 mM NaCl, 0.08% $C_{10}E_5$ or 0.03% DDM (Ton subcomplex); or 20 mM HEPES-NaOH, pH 7.0, 150 mM NaCl, and 0.05% DDM (Ton complex).

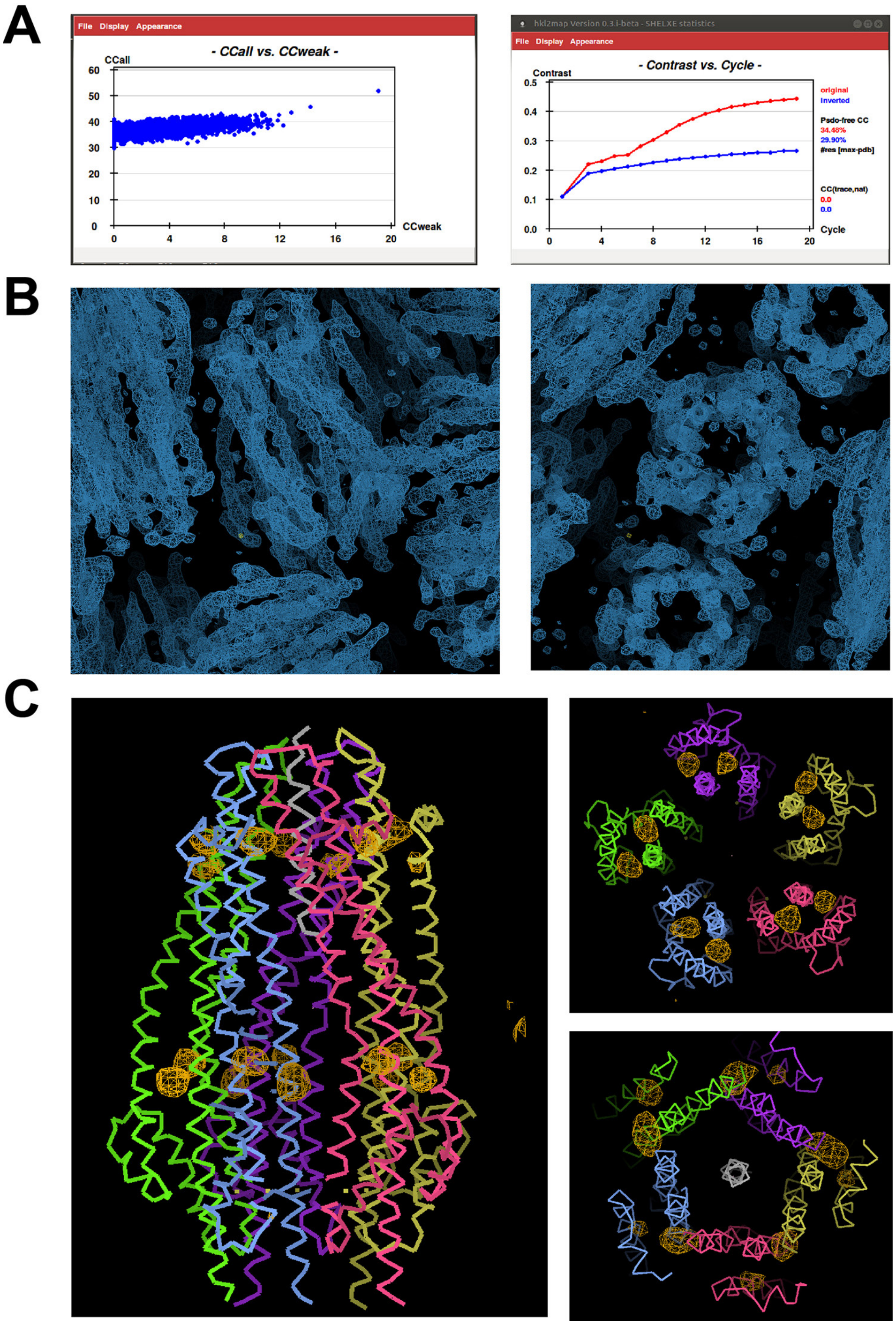
For DEER measurements, the samples were diluted with D_2O to a final concentration of 30% and cryoprotected with 10% v/v D8-glycerol before being flash frozen in liquid nitrogen. Continuous wave (CW) electron paramagnetic resonance (EPR) experiments were carried out at room temperature on a bench-top X-band MiniScope MS 400 (Magnettech by Freiberg Instrument) at 9.5 GHz (X-band) with 2.5 mW microwave power, 15 mT sweep width and 0.15 mT modulation amplitude. Spin labelling efficiency was calculated from the second integral of the derivative spectra compared to a standard spin concentration of 100 μM (Tempol in water). The ExbB native cysteine C25 was labelled with a 50% efficiency, while the ExbD mutants were labelled with efficiencies >80%. DEER measurements were initially performed at ETH Zurich on a commercial Bruker ELEXSYS-II E580 Q-band spectrometer (34–35 GHz) and later on a Bruker ELEXSYS E580Q-AWG dedicated pulse Q-band spectrometer operating at 34–35 GHz. Both spectrometers were equipped with a TWT amplifier (150 W) and a home-made rectangular resonator (from ETH Zurich) enabling the insertion of 30–40 μl sample volume in quartz tubes with 3 mm outer diameter[51]. Dipolar time evolution data were acquired using the four-pulse DEER experiment at 50 K. All pulses were set to be rectangular with 12 ns length, with the pump frequency at the maximum of the echo-detected field swept spectrum, 100 MHz higher than the observer frequency. Deuterium nuclear modulations were averaged by increasing the first interpulse delay by 16 ns for 8 steps as previously described[51]. The background of the normalized DEER primary data ($V(t)/V(0)$) was fitted with optimized dimensions from 2.5 to 3.2 and the resulting normalized secondary data ($F(t)/F(0)$) were converted by model-free Tikhonov regularization to distance distributions with the software DeerAnalysis2015[52,53]. The simulation of the possible spin label rotamers populated at selected positions in the protein was performed using the Matlab program package MMM2015.1 using the MTSL ambient temperature library[54].

Crosslinking. The $ExbB_{C25S}$–$ExbD_{E113C}$ complex ($ExbD_{E113C}$ was in the pACYC vector containing an N-terminal strep-tag and a C-terminal 6×HIS tag) was expressed and purified as described earlier. To prepare the sample for crosslinking, the sample was incubated at 4 °C with 5 mM DTT for at least 1 h. The DTT was then removed using a desalting column in 20 mM HEPES, pH 7.0, 150 mM NaCl, and 0.1% DDM. The crosslinker 1,8-bismaleimidodiethylenglycol ($BM(PEG)_2$) (Pierce) was added at a final concentration of 0.2 mM and the reaction was incubated at 4 °C overnight. The sample was concentrated and passed over a Superose 6HR 10/30 gel filtration column using 20 mM HEPES-NaOH, pH 7.0, 150 mM NaCl, and 0.035% DDM on an AkTA Purifier system (GE Healthcare). The results were visualized by SDS–PAGE analysis.

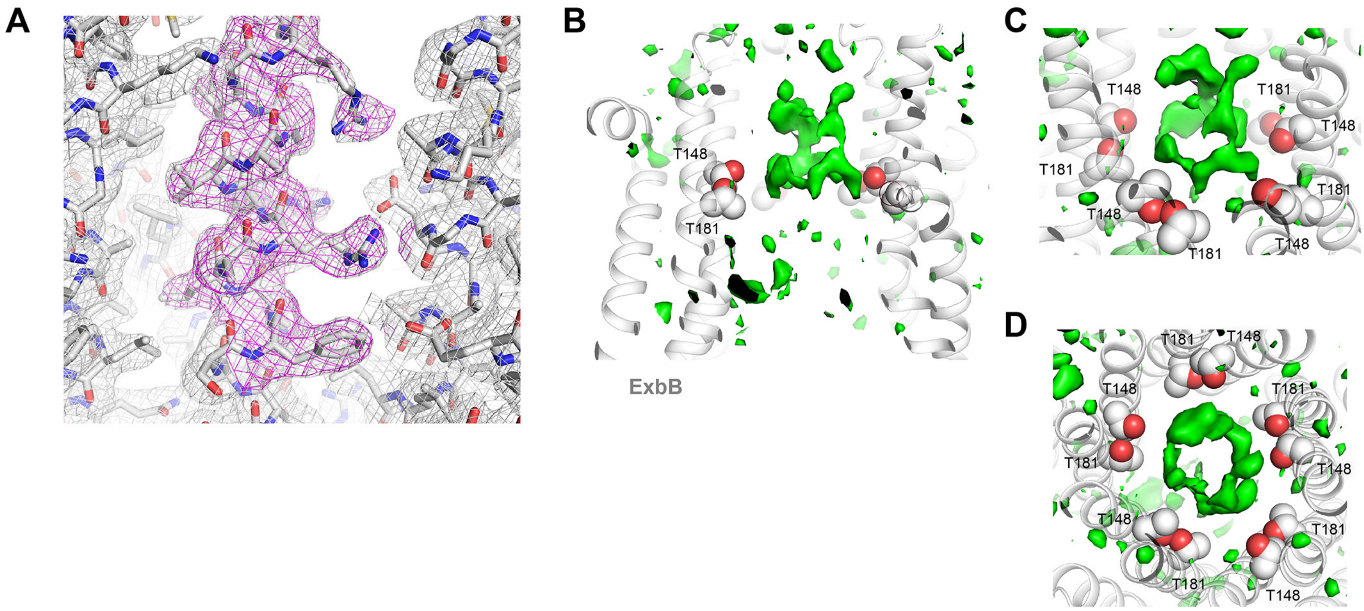
Reconstitution in liposomes. Protein complexes were reconstituted into liposomes by dialysis of the protein–lipid–detergent mixture. Lipids (DOPG, DOPC and DOPE) dissolved in chloroform were mixed in a molar ratio of 2:3:5. Chloroform was removed by vortexing in a stream of nitrogen gas in a glass tube followed by drying in vacuum for 2–3 h. The lipid film was hydrated in 1 ml TN buffer (10 mM Tris-HCl, pH 7.5, 50 mM NaCl), followed by five cycles of freeze–thaw and sonication using a water bath sonicator until the suspension of lipids became clear (10–15 min). For proteoliposome preparation, small unilamellar vesicles (SUVs) were mixed with octylglucoside (final concentration, 2%) and then proteins added to achieve a molar ratio of total lipid to protein ∼500–2,000 mol/mol. After 1 h incubation in ice, the lipid–protein–detergent mixture was dialysed into 10 mM Tris-HCl, pH 7.5, 0.3 M sucrose, and 50 mM KCl for 30–40 h using a dialysis membrane with a MW cut-off pore size of 10 kDa.

Planar-lipid bilayer measurement of ion-conduction. Mueller-Rudin type planar bilayer membranes were formed on a 0.2-mm diameter aperture in a partition that separates two 1-ml compartments, using a mixture of lipids, DOPG, DOPC and DOPE, at a molar ratio of 2:3:5 (10 mg/ml) in n-decane, applied by a brush technique[55]. The aqueous solution in both compartments consisted of 2 mM KP_i, pH 7.0, and 0.1 M and 0.4 M KCl in the *cis*- and *trans*-compartments, respectively. To study the pH dependence of channel activity, bathing solutions were buffered with 2 mM Na-acetate (pK 4.8), Na-cacodylate (pK 6.2), and Tris (pK 8.3). The pH of the bathing solution was changed by adding 10–20 μl 0.1 M HCl or KOH. The *cis*-side of the planar bilayer is defined as that to which the electrical potential is applied. Proteoliposomes, 0.1–2 μl, were added to the *trans*-compartment, and the solutions were stirred until the transmembrane current appeared. A large concentration of an osmolyte inside of the liposomes and the transmembrane KCl concentration gradient caused proteoliposome fusion with the pre-formed planar lipid membrane bilayer. The transmembrane current was measured in voltage-clamp mode with Ag/AgCl electrodes and agar bridges, using a BC-525C amplifier (Warner Instruments). The single-channel conductance of the ExbB–ExbD complexes was measured in symmetrical salt conditions: 0.1 M KCl solution, pH 7.5, at a holding potential of +50 or −50 mV. For ion selectivity experiments, zero-current potential (V_{rev}) was determined from volt-ampere characteristics measured in asymmetric salt conditions. Relative cation/anion permeability was calculated using the Goldman-Hodgkin-Katz equation[56].

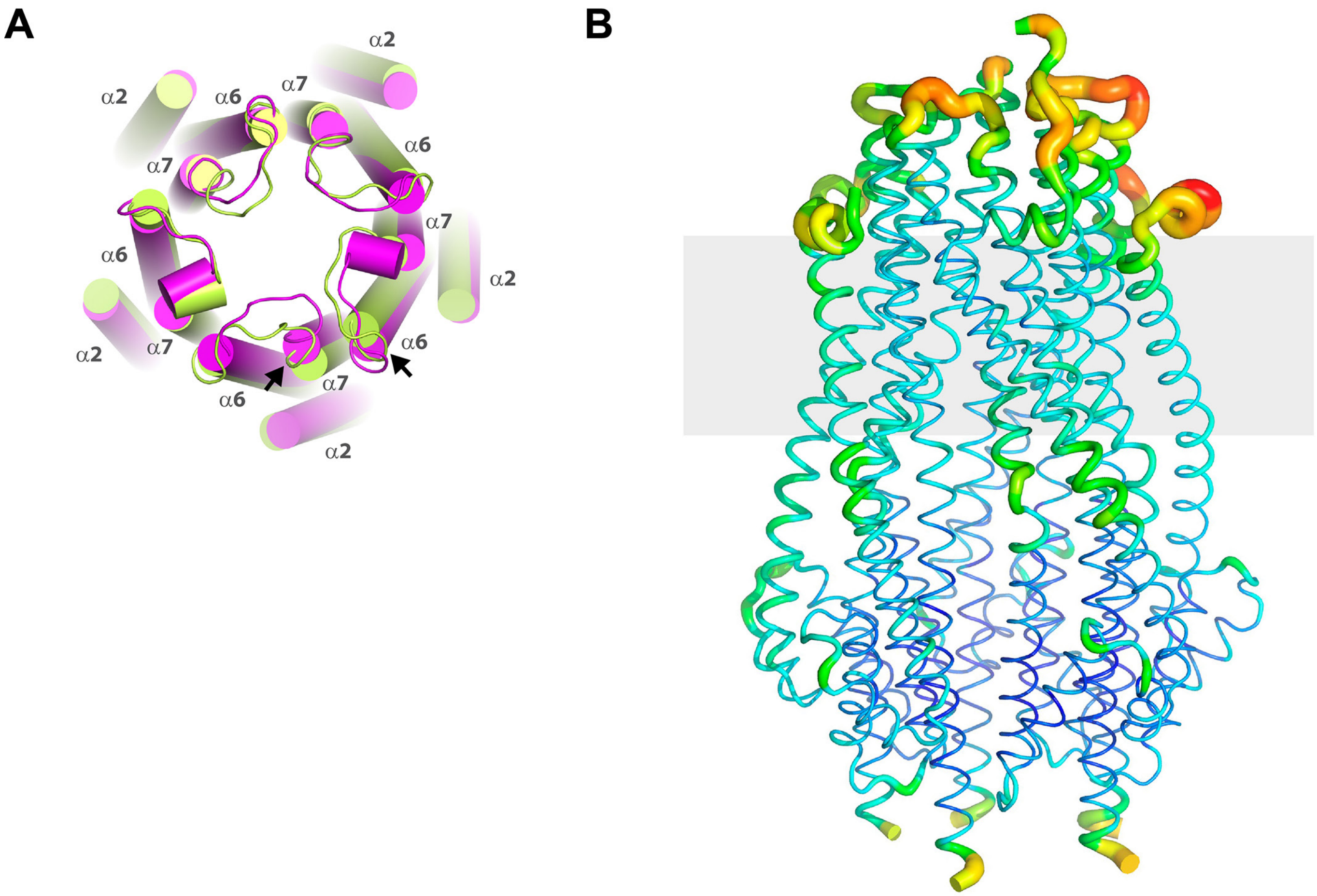
41. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9,** 671–675 (2012).
42. Whitmore, L. & Wallace, B. A. Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases. *Biopolymers* **89,** 392–400 (2008).
43. Minor, Z. O. W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276,** 307–326 (1997).
44. Winter, G. xia2: an expert system for macromolecular crystallography data reduction. *J. Appl. Crystallogr.* **43,** 186–190 (2010).
45. Schneider, T. P. T. R. HKL2MAP: a graphical user interface for phasing with SHELX programs. *J. Appl. Crystallogr.* **37,** 843–844 (2004).
46. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66,** 213–221 (2010).
47. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40,** 658–674 (2007).
48. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372,** 774–797 (2007).
49. Pettersen, E. F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25,** 1605–1612 (2004).
50. Schenk, A. D., Philippsen, A., Engel, A. & Walz, T. A pipeline for comprehensive and automated processing of electron diffraction data in IPLT. *J. Struct. Biol.* **182,** 173–185 (2013).
51. Polyhach, Y. *et al.* High sensitivity and versatility of the DEER experiment on nitroxide radical pairs at Q-band frequencies. *Phys. Chem. Chem. Phys.* **14,** 10762–10773 (2012).
52. Jeschke, G. *et al.* DeerAnalysis2006—a comprehensive software package for analyzing pulsed ELDOR data. *Appl. Magn. Reson.* **30,** 473–498 (2006).
53. von Hagens, T., Polyhach, Y., Sajid, M., Godt, A. & Jeschke, G. Suppression of ghost distances in multiple-spin double electron-electron resonance. *Phys. Chem. Chem. Phys.* **15,** 5854–5866 (2013).
54. Polyhach, Y., Bordignon, E. & Jeschke, G. Rotamer libraries of spin labelled cysteines for protein studies. *Phys. Chem. Chem. Phys.* **13,** 2356–2366 (2011).
55. Mueller, P., Rudin, D. O., Tien, H. T. & Wescott, W. C. Reconstitution of cell membrane structure in vitro and its transformation into an excitable system. *Nature* **194,** 979–980 (1962).
56. Hille, B. *Ion Channels of Excitable Membranes* 3rd edn (Sinauer, 2001).



Extended Data Figure 1 | Structure determination for the Ton subcomplex (ExbB–ExbD$_{\Delta peri}$) using Se-SAD at 5.2 Å resolution. **a**, The initial structure of the Ton subcomplex was solved by Se-SAD using anisotropic data extending to 5.2 Å resolution. The data from two crystals were processed with Xia2 and the initial sites found using HKL2MAP v0.3, which found a single solution every ~10,000 tries; resolution limits were also important for finding a solution. **b**, The sites were then input into AUTOSOL/PHENIX for site refinement and density modification, producing density maps (blue mesh) which clearly showed five-fold symmetry and allowed an initial model of a monomer to be built, consisting almost entirely of α-helices. This model was then used as a search model for molecular replacement to solve the 2.6 Å native structure (data obtained from a single crystal). **c**, Anomalous different map (orange mesh) showing density for the selenium sites in the 5.2 Å Se-incorporated structure.

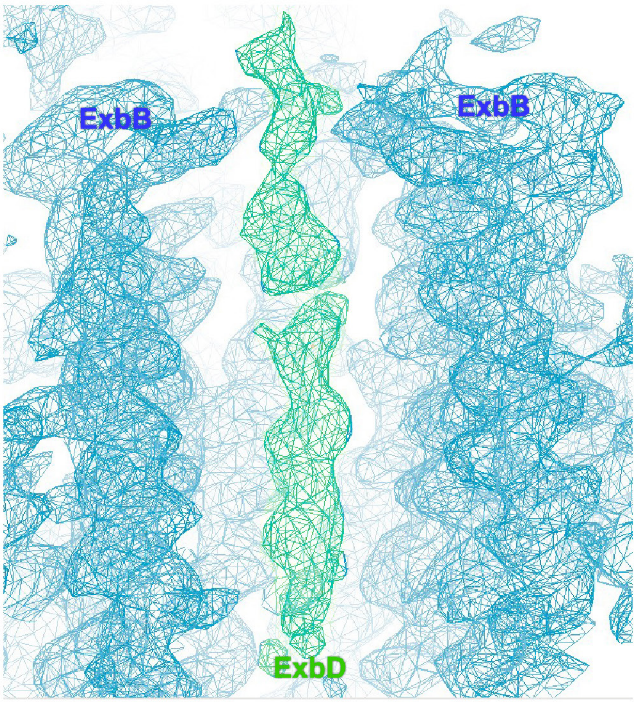


Extended Data Figure 2 | Representative electron density for the native Ton subcomplex (ExbB–ExbD$_{\Delta\text{peri}}$) solved to 2.6 Å resolution. **a**, Representative electron density map ($2F_o-F_c$ contoured at 1.0σ, grey mesh; $2F_o-F_c$ omit map (omitting residues 113–124) contoured at 1.0σ, magenta mesh) along residues 113–124 within helix α5. **b**, Cutaway view of the transmembrane pore of ExbB (grey ribbon) from the native structure at pH 7.0 showing ring-like difference density (green isosurface, F_o-F_c map contoured at 2.5σ) along the conserved residues T148 and T181 (grey and red spheres). **c**, **d**, Tilted view (**c**) and an orthogonal view (**d**) (relative to **a**) of the ring-like density. Structures were determined using data obtained from a single crystal in each case.

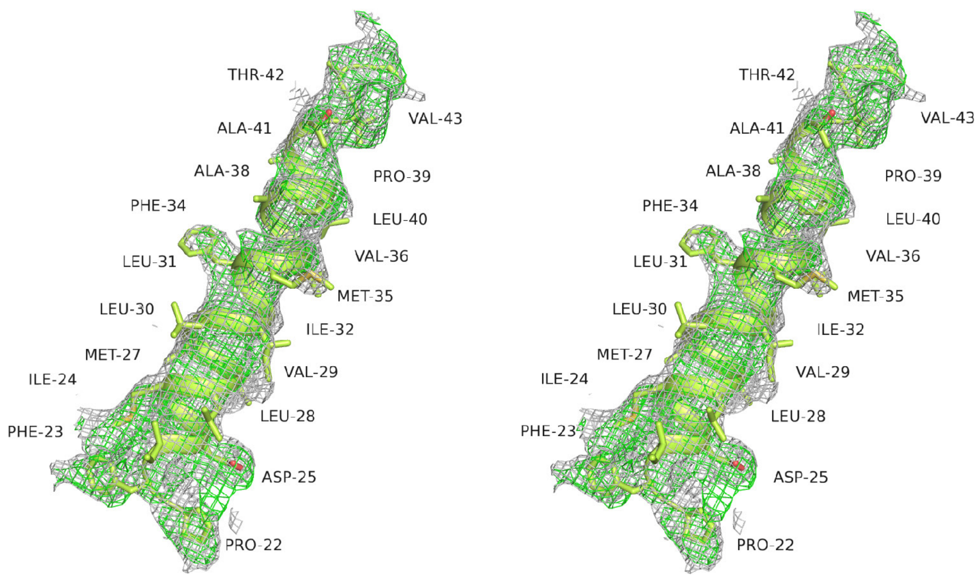


Extended Data Figure 3 | Helical shifts and overall flexibility in the ExbB pentamer. a, Two pentamers were observed per asymmetric unit within the crystal structure. Shown here is pentamer 1 (green) aligned with pentamer 2 (magenta), illustrating slight shifts in a number of the helices (cylinders) between the two pentamers, with the largest shifts indicated by black arrows. The loops connecting α6 and α7 also show variability between monomers and pentamers. **b**, The TonB subcomplex (ExbB–ExbD$_{\Delta peri}$) showing a B-factor putty representation with values ranging from the most ordered in blue to the most disordered in red.

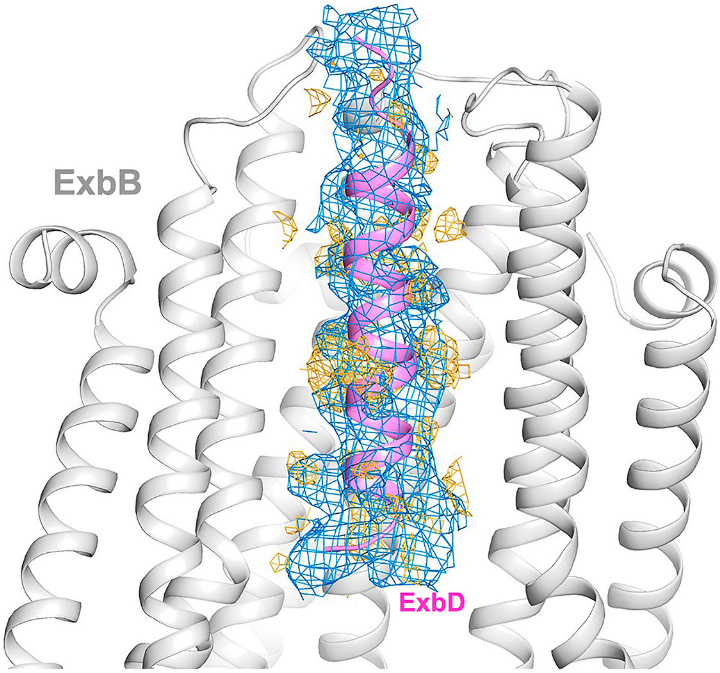
A



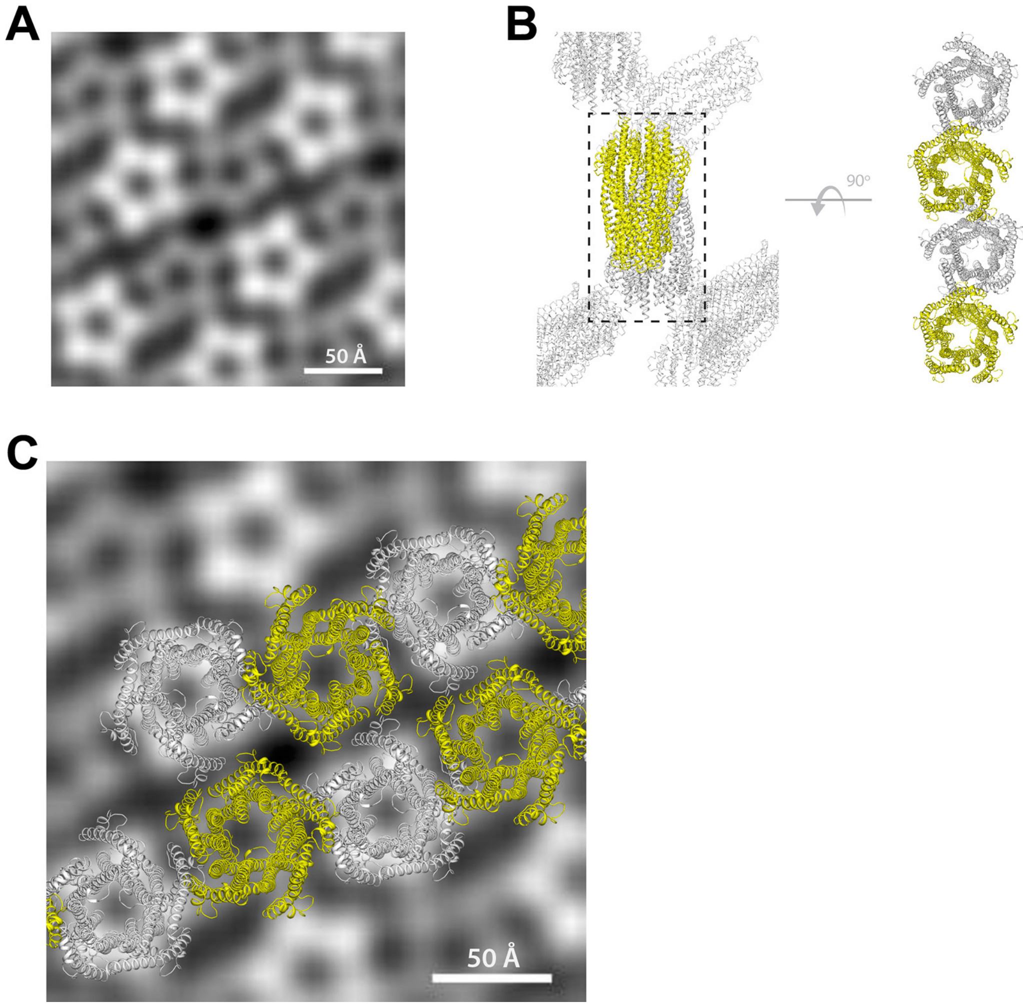
B



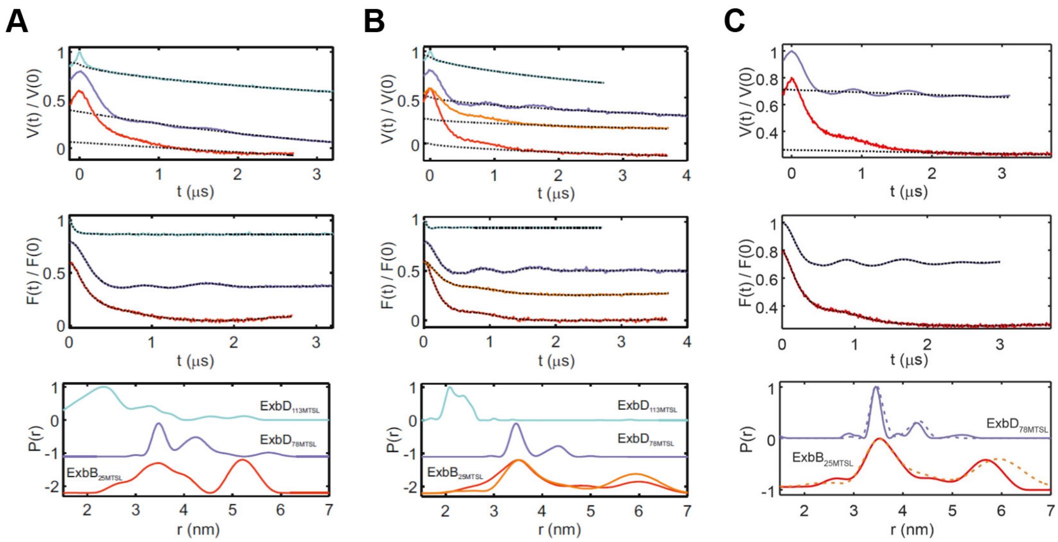
Extended Data Figure 4 | Electron density for the transmembrane helix of ExbD. a, Omit map ($2F_o-F_c$, contoured at 1.0σ) along the transmembrane pore of ExbB. The density corresponding to the ExbB pentamer is shown in blue mesh, while the density corresponding to the transmembrane helix of ExbD is shown in green mesh. **b**, Stereoimage showing the density ($2F_o-F_c$, contoured at 0.8σ, grey mesh; $2F_o-F_c$ omit map (omitting the transmembrane helix of ExbD), contoured at 0.8σ, green mesh) for the transmembrane helix of ExbD after building and refinement.



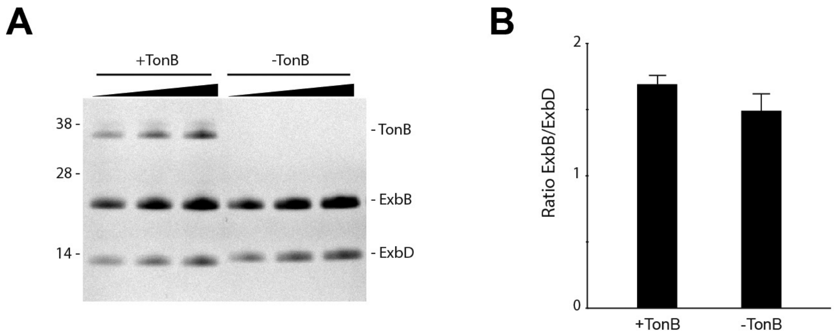
Extended Data Figure 5 | Comparison of observed density for crystal structures of ExbB–ExbD$_{\Delta peri}$ solved at pH 7.0 versus pH 4.5. The presence of electron density for the transmembrane helix of ExbD (magenta ribbon) was dependent on the pH at which the crystals were grown. At pH 7.0, we observed little density (orange mesh) inside the transmembrane pore of the ExbB (grey ribbon) pentamer (see also Extended Data Fig. 3). However, for the structures solved at pH 4.5, we observed clear density (blue mesh) for the transmembrane helix of ExbD, albeit to varying degrees. Density maps ($2F_o - F_c$) are contoured at 1.0σ.



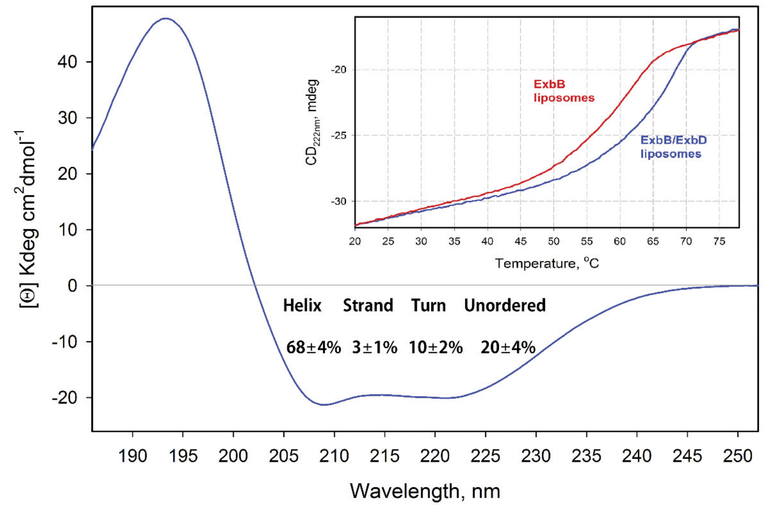
Extended Data Figure 6 | Packing similarities of the 2D and 3D crystals used for electron microscopy and X-ray crystallography. a, Averaged projection map from the electron microscopy analysis on 2D crystals. Five images were analysed, and a representative averaged projection map was calculated from 900 sub-images. The averaged map shows two different populations of the pentamer that are similar in size but differ in level intensity owing to opposite orientations of the complex within the crystal; a similar packing arrangement was also observed in our crystal structures. ExbD was not detected in our electron microscopy studies, probably owing to disorder of the globular domain, which is anchored to the membrane by a long unstructured linker[15]. **b**, Packing of the complex in the X-ray crystal structure from 3D crystals. The right side indicates an orthogonal view highlighting a single row of molecules from the lattice (black dashed box). **c**, Fitting the row of molecules from the 3D lattice (X-ray) from **b** onto the averaged projection map from the 2D crystals (electron microscopy) to highlight the consistency observed in packing.



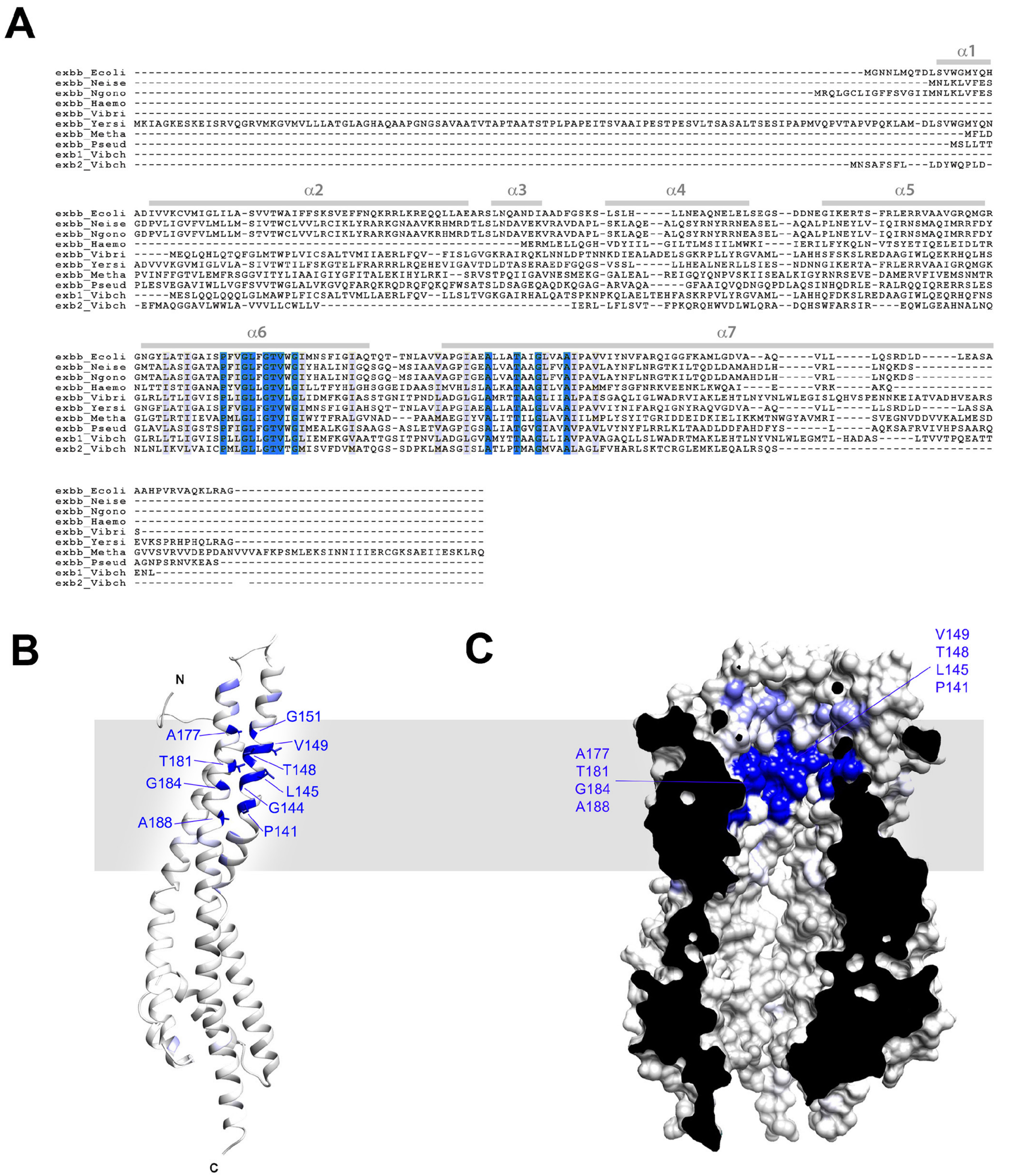
Extended Data Figure 7 | DEER traces and analysis. Ton subcomplex ($ExbB_{C25}$–ExbD, $ExbB_{C25S}$–$ExbD_{N78C}$, and $ExbB_{C25S}$–$ExbD_{E113C}$) in 0.08% $C_{10}E_5$ (**a**) and in 0.03% DDM (**b**), and the fully assembled Ton complex ($TonB_{C18A}$–$ExbB_{C25}$–ExbD and $TonB_{C18A}$–$ExbB_{C25S}$–$ExbD_{N78C}$) in 0.05% DDM (**c**). Upper panels, experimental Q-band DEER primary data $V(t)/V(0)$ (coloured lines, cyan $ExbD_{113MTSL}$; violet $ExbD_{78MTSL}$; red and orange, $ExbB_{25MTSL}$) and simulated background functions (dotted line). Middle panels, DEER traces after background correction (coloured lines) and fit with DeerAnalysis2015 (dotted lines) with Tikhonov regularization parameters from 10 to 100 adjusted via L-curve analysis and data validation. Lower panels, obtained distance distributions. For the pentameric ExbB sample (50% labelling efficiency), a modulation depth >0.45 was obtained, indicating the presence of a multi-spin system. For the sample solubilized in DDM, longer DEER traces were obtained (4 μs) to better characterize the long distance peak of 5–6 nm in $ExbB_{25MTSL}$. Additionally, for all panels, another DEER trace was measured after decreasing the microwave power of the 12-ns pump pulse to 25% (orange line) to suppress ghost peaks arising from the presence of more than two spins in the system. The resulting distance distribution (orange) was found to be very similar to that obtained with 100% microwave power (red), showing that no ghost peak artefacts were present. The lower modulation depth observed for the ExbD samples labelled at position 113 with respect to those labelled at position 78 (both labelling efficiency >80%) may be due to the presence of distances <1.5 nm (predicted by the simulations), which are outside of the sensitivity range of the technique, or to destabilization of the ExbD dimer induced by the label. The bottom of **c** shows a comparison of the Ton subcomplex in DDM (dashed lines from **b**) to the fully assembled Ton complex (solid lines). All panels show data from single experiments.



Extended Data Figure 8 | Densitometry of the purified fully assembled Ton complex. a, SDS–PAGE gel of the Ton complex (+TonB) and the Ton subcomplex (−TonB) at increasing concentrations. **b**, Bar graph showing the comparison of the ExbB–ExbD ratio within the Ton complex (+TonB) and the Ton subcomplex (−TonB) indicating that association of TonB with the Ton subcomplex does not change the stoichiometric ratio of the components. While we see a slight difference in the ExbB–ExbD ratio values in the presence or absence of TonB, the observed difference is too small to suggest an altered stoichiometry between ExbB and ExbD. Three representative lanes for each sample are shown in **a**; however, five lanes were used for all calculations. Densitometry analysis was performed with ImageJ and mean values and standard errors calculated using Microsoft Excel. For purifications of the Ton complex (+TonB), five purification experiments were performed and one representative is shown. For purifications of the Ton subcomplex (−TonB), ∼50 purifications were performed and one representative is shown.



Extended Data Figure 9 | Circular dichroism analysis of secondary structure and thermal stability of the Ton subcomplex. Far-UV circular dichroism spectrum (185–260 nm) of the Ton subcomplex (ExbB–ExbD) with the calculated percentage of secondary structure shown. Contents of regular and distorted α-helical structures, 47 and 21%, respectively, were combined during the calculation of secondary structure contributions. Inset, comparison of the thermal stability of the Ton subcomplex (blue) versus ExbB alone (red) measured through the temperature dependence of the circular dichroism signal amplitude at 222 nm. Both panels show data from a single experiment.



Extended Data Figure 10 | Sequence conservation of ExbB orthologues mapped onto the crystal structure. a, Clustal W alignment of ExbB sequences from: *E.coli* K12 (P0ABU7), *Neisseria meningitidis* (P64100), *Neisseria gonorrhoeae* (Q5F711), *Haemophilus ducreyi* (O51808), *Vibrio harveyi* (D0XEN5), *Yersinia pestis* (D1TTA4), *Methanothermobacter thermautotrophicus* (O27101), *Pseudomonas aeruginosa* (G3XCW0), ExbB1 of *Vibrio cholerae* (O52043) and ExbB2 of *Vibrio cholerae* (AAC69454). **b**, Conservation mapped onto the ExbB structure with Chimera. The most conserved residues are in blue and found in α6 (TM2) and α7 (TM3) of the ExbB structure. An extensive alignment that also includes sequences from the Tol and Mot systems shows similar results[22]. **c**, Cutaway molecular surface of ExbB pentamer with the most conserved residues mapped onto the surface.

doi:10.1038/nature19367

X-ray structures define human P2X$_3$ receptor gating cycle and antagonist action

Steven E. Mansoor[1,2], Wei Lü[1], Wout Oosterheert[1]†, Mrinal Shekhar[3], Emad Tajkhorshid[3] & Eric Gouaux[1,4]

P2X receptors are trimeric, non-selective cation channels activated by ATP that have important roles in the cardiovascular, neuronal and immune systems. Despite their central function in human physiology and although they are potential targets of therapeutic agents, there are no structures of human P2X receptors. The mechanisms of receptor desensitization and ion permeation, principles of antagonism, and complete structures of the pore-forming transmembrane domains of these receptors remain unclear. Here we report X-ray crystal structures of the human P2X$_3$ receptor in apo/resting, agonist-bound/open-pore, agonist-bound/closed-pore/desensitized and antagonist-bound/closed states. The open state structure harbours an intracellular motif we term the 'cytoplasmic cap', which stabilizes the open state of the ion channel pore and creates lateral, phospholipid-lined cytoplasmic fenestrations for water and ion egress. The competitive antagonists TNP-ATP and A-317491 stabilize the apo/resting state and reveal the interactions responsible for competitive inhibition. These structures illuminate the conformational rearrangements that underlie P2X receptor gating and provide a foundation for the development of new pharmacological agents.

Integral membrane proteins that recognize extracellular nucleotides were defined in 1976 and termed purinergic receptors[1–3]. Two families of purinergic receptors have since been established: ligand-gated P2X receptor ion channels[4] and G-protein coupled P2Y receptors[5]. P2X receptors are found throughout eukaryotes[6]; in humans, they are expressed in a wide variety of cells and modulate processes as diverse as platelet activation, smooth muscle contraction, synaptic transmission, nociception, inflammation, hearing and taste[7,8], making P2X receptors important pharmacological targets[9].

The seven mammalian P2X receptor subtypes, denoted P2X$_1$–P2X$_7$, form homotrimeric and heterotrimeric complexes[4,10,11]. All subunits share a common topology containing intracellular termini, two transmembrane helices forming the ion channel, and a large extracellular domain containing the orthosteric ATP binding site[11,12]. Whereas all P2X receptors are non-selective cation channels that are permeable to Na^+ and Ca^{2+} and activated by ATP[13], the pharmacology of receptor subtypes varies with respect to sensitivity to ATP analogue agonists and to small molecule antagonists. Thus, although 2′-3′-O-(2,4,6,-trinitrophenyl) adenosine 5′-triphosphate (TNP-ATP) is the prototypical nanomolar-affinity antagonist of P2X$_{1,3}$ receptors, it binds 1,000-fold less tightly to P2X$_{2,4,7}$ receptors[9,14]. The kinetics of ion channel gating also vary by subtype, with P2X$_{2,4,5,7}$ receptors showing slow and incomplete desensitization and P2X$_{1,3}$ receptors undergoing rapid and nearly complete desensitization[15,16].

Membrane-proximal regions within the cytoplasmic termini play important roles in receptor desensitization[17–25], but the detailed molecular mechanism of desensitization is unknown. Mechanisms that have been proposed are similar to the 'hinged lid' or 'ball and chain' models described for voltage-gated sodium and shaker potassium channels, respectively, with a distinct but unidentified desensitization gate[21,26]. No structure of a P2X receptor in the desensitized state has been published, to our knowledge, and currently available structures of the zebrafish P2X$_4$ receptor (zfP2X$_4$) in the apo and open state conformations do not visualize cytoplasmic residues[27–29]. There is also concern that the available structure of zfP2X$_4$ bound to ATP[27] may not represent a physiological state because the truncated crystallization construct, which lacks both terminal domains, might distort the pore architecture[12,30–32]. The mechanisms by which antagonists inhibit ion flow through P2X receptors remain elusive. A recent NMR study suggested that TNP-ATP inhibits activation by closing the extracellular fenestrations to ion access, rather than by stabilizing a closed-pore conformation[33]. To understand the molecular mechanisms of activation and antagonism of P2X receptors, we crystallized the human P2X$_3$ (hP2X$_3$) receptor in an apo/resting state, an agonist-bound/open-pore state, an agonist-bound/closed-pore/desensitized state, and two antagonist-bound/closed states.

Crystallization and structure determination

The hP2X$_3$ crystallization construct spans residues D6–T364 and is defined as hP2X$_3$-MFC. It binds ATP with a dissociation constant (K_d) of 2.8 nM and has wild-type gating properties, as shown by scintillation proximity assays (SPA)[34] and two-electrode voltage clamp (TEVC; Extended Data Fig. 1a, b), respectively. Notably, hP2X$_3$-MFC demonstrates fast desensitization kinetics, the hallmark of homotrimeric P2X$_3$ receptors[35,36]. Three rat P2X$_2$-specific amino acid substitutions[21] were made at homologous residues in the N terminus of hP2X$_3$ to generate hP2X$_3$-MFC-T13P/S15V/V16I (or hP2X$_3$-MFC$_{slow}$), a construct with similar affinity (K_d = 3.3 nM) for ATP (Extended Data Fig. 1c) but with slow and incomplete desensitization (Extended Data Fig. 1d). The structure of the ATP-bound/open-pore state (Fig. 1a–c) was obtained using hP2X$_3$-MFC$_{slow}$ whereas hP2X$_3$-MFC was used to determine the structure of the ATP-bound/closed-pore/desensitized state (Fig. 1d–f).

We further crystallized hP2X$_3$-MFC$_{slow}$ in an apo/resting state (Fig. 1g–i) and in complex with two high-affinity P2X$_3$ competitive antagonists (TNP-ATP[14,37] and A-317491 (ref. 38)). Both antagonists inhibited ATP-induced currents from hP2X$_3$-MFC and hP2X$_3$-MFC$_{slow}$ expressed in oocytes, and TNP-ATP displaced radioactive

[1]Vollum Institute, Oregon Health & Science University, Portland, Oregon 97239, USA. [2]Knight Cardiovascular Institute, Oregon Health & Science University, Portland, Oregon 97239, USA. [3]Department of Biochemistry, Center for Biophysics and Quantitative Biology, and Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA. [4]Howard Hughes Medical Institute, Oregon Health & Science University, Portland, Oregon 97239, USA. †Present address: Crystal and Structural Chemistry, Bijvoet Center for Biomolecular Research, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands.

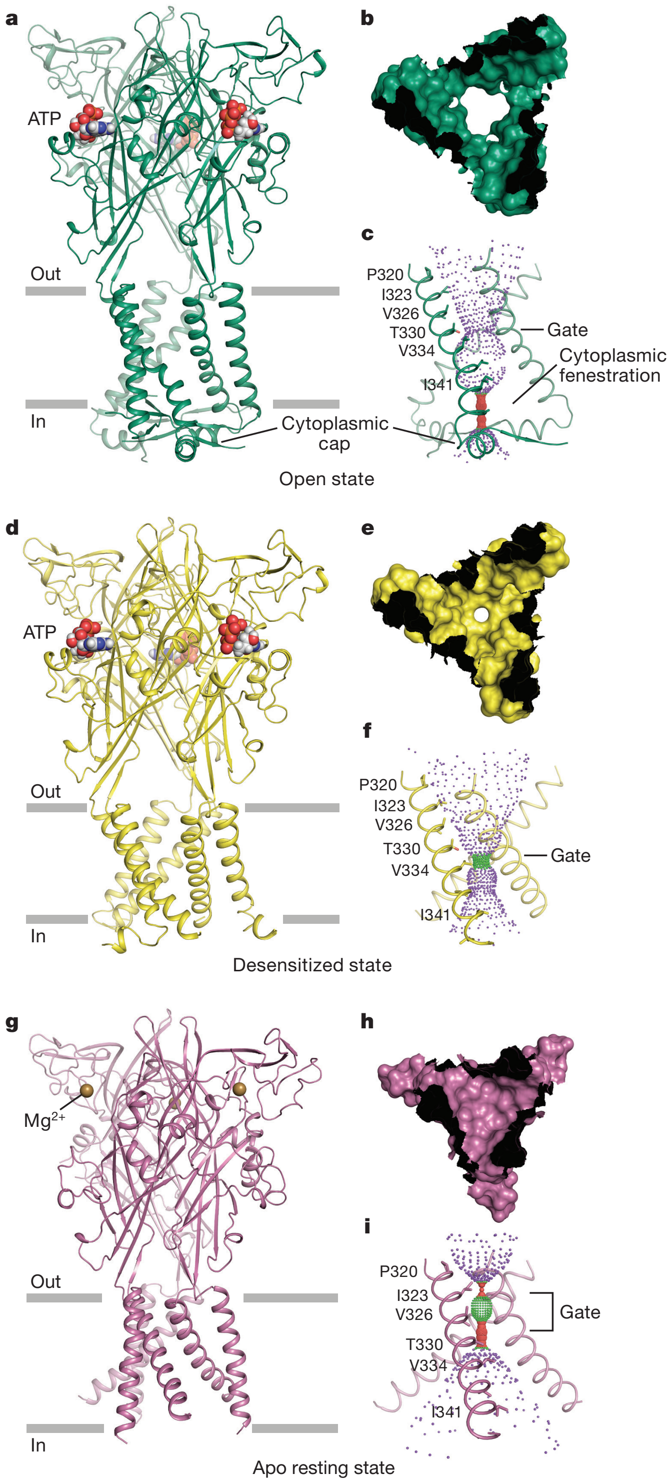


Figure 1 | Architecture and pore structure for major conformational states of the gating cycle of hP2X$_3$. a–i, Cartoon representation of each hP2X$_3$ structure shown parallel to the membrane as a side view, perpendicular to the membrane from the extracellular side as a surface representation, and the ion permeation pathway, respectively, are drawn for open state (**a**–**c**), desensitized state (**d**–**f**), and apo state (**g**–**i**). Each conformational state is colour-coded unless otherwise noted: open state in green, desensitized state in yellow, and apo state in red-purple. For the pore size plots, different colours represent different radii, as calculated by the program HOLE: red <1.15 Å, green 1.15–2.30 Å, and purple >2.30 Å.

ATP from detergent-solubilized hP2X$_3$-MFC$_{slow}$ (Extended Data Fig. 1e–g). Prolonged application of ATP to oocytes expressing hP2X$_3$-MFC$_{slow}$ resulted in a residual current that was blocked by the competitive antagonists, suggesting that a fraction of these receptors did not desensitize (Extended Data Fig. 1h, i). The hP2X$_3$ structures were refined to good crystallographic statistics and stereochemistry (Extended Data Table 1).

Overall architecture

The hP2X$_3$ structures follow the iconic trimeric P2X receptor architecture[27–29], possessing a large hydrophilic extracellular domain, six α-helices forming the transmembrane domain, and intracellular termini (Fig. 1). The shape of each protomer resembles that of a dolphin[27,29] (Extended Data Fig. 2a). The open state structure of hP2X$_3$ contains ATP in the ligand-binding pocket and an open pore (Fig. 1a–c) whereas the desensitized state structure has ATP in the pocket but a closed pore (Fig. 1d–f). Although the extracellular domains and binding pockets of the desensitized and open states of hP2X$_3$ are similar, there are striking differences in the transmembrane domains and at the gates (Extended Data Fig. 3). Both hP2X$_3$ structures have transmembrane domains that are of sufficient length to cross a lipid bilayer, and the desensitized state structure has a pore architecture not previously observed for any P2X structure. The open state structure of hP2X$_3$ visualizes cytoplasmic residues that were truncated in the open state structure of zfP2X$_4$ (ref. 27) (Extended Data Fig. 2b) and forms a domain termed the 'cytoplasmic cap' (Fig. 1a, c).

The apo structure of hP2X$_3$ has an empty ligand-binding pocket and a closed pore (Fig. 1g–i). An alignment to the apo structure of zfP2X$_4$ (ref. 29) reveals several unique features of the hP2X$_3$ apo structure, including a more complete transmembrane domain, different residues defining the pore constriction, and a Mg^{2+} ion bound in the head domain (Extended Data Fig. 2c). Comparison between the hP2X$_3$ structures and previously published zfP2X$_4$ structures emphasizes the longer transmembrane domains and the cytoplasmic domain of hP2X$_3$ (Extended Data Fig. 2b–d). In the antagonist-bound state, the competitive antagonists TNP-ATP and A-317491 occupy the orthosteric ligand-binding pocket and the ion channel pore is closed and nearly identical to the apo/resting state (Extended Data Fig. 4).

Ion channel pore

To determine the functional state of each hP2X$_3$ structure, we analysed the conformation of the ion channel pore, together with alterations in the size and shape of cavities, vestibules and fenestrations throughout the receptor (Extended Data Fig. 3a, b). Transmembrane helix 2 (TM2) lines the pore lumen, with residues I323, V326, T330, and V334 facing the pore[27,29,39] (Fig. 1c, f, i). I323 defines the extracellular boundary of the gate in the apo state (pore radius 0.3 Å), whereas T330 defines the cytoplasmic boundary of the gate (pore radius 0.7 Å) (Extended Data Fig. 3b, c). A third residue, V326, also contributes to the pore occlusion. These openings are too narrow to pass dehydrated Na^+ ions[40] and define the ion channel as closed (Fig. 1h, i). The solvent-accessible surface, dimensions, and residues lining the gate for both antagonist-bound structures are similar to those of the apo state structure, demonstrating that these competitive antagonists stabilize an apo/resting-like state of the receptor (Extended Data Fig. 4c–f).

The open state structure of hP2X$_3$ has a continuous pore through the transmembrane domain with a minimum radius of 3.2 Å (Extended Data Fig. 3c), which is large enough to pass partially hydrated Na^+ ions[41] and defines the ion channel gate as open (Fig. 1b, c). Compared to the apo structure, I323 and V326 in the open state structure have been translated upward towards the extracellular surface and rotated outward, away from the pore's centre, to open the pore. T330, which defined the cytoplasmic boundary of the closed gate in the apo state, now defines the narrowest region of the pore in the open state. In hP2X$_3$, T330 and S331 are the only hydrophilic residues lining the middle of the pore. For the rat P2X$_2$ receptor, a threonine residue at the equivalent position to T330 of hP2X$_3$ has been implicated in ion selectivity[42], suggesting that T330 might interact with permeating cations.

A single residue, V334, defines the constriction site of the desensitized state with a pore radius of 1.5 Å, too narrow to pass hydrated Na^+ ions (Fig. 1e, f and Extended Data Fig. 3c). From the open to the desensitized states, V334 translates upward towards the extracellular surface and rotates inward to block the pore. To ensure the density in the pocket was truly ATP, we soaked the crystals with an ATP derivative, the $P2X_3$ agonist 2-(methylthio)adenosine 5′-triphosphate (2-methylthio-ATP), and collected native and anomalous sulfur diffraction data (Extended Data Tables 1, 2). These soaked crystals retained the same receptor and pore structures but had a density in the binding pocket consistent with 2-methylthio-ATP, confirmed through the anomalous signal of the sulfur atom, providing evidence that the structure represents an agonist-bound/closed-pore/desensitized state (Extended Data Fig. 5).

Channel opening

Comparing the apo and open state structures of $hP2X_3$ demonstrates the extensive structural differences between these two conformational states, emphasizing the role of the 'cytoplasmic cap' in stabilizing the open state (Fig. 2a, b). The cytoplasmic cap includes elements of secondary structure from both termini, including two sequential β-strands from the N terminus and a β-strand from the C terminus (Fig. 2c and Extended Data Fig. 2a). The tertiary structure of the cytoplasmic cap is defined by a network of three β-sheets that sit beneath the transmembrane domain, capping the cytoplasmic surface of the pore. The C-terminal β-strand of each protomer interacts with the N-terminal β-strands of each of the other two protomers to form a small β-sheet (Fig. 2c, d). Each of the three β-sheets incorporates one β-strand from each of the three protomers, illustrating how domain swapping knits receptor subunits together on the cytoplasmic side of the membrane. The cytoplasmic cap is observed only in the ATP-bound open state structure, suggesting that the cap-forming elements are flexible and disordered in the apo state. Indeed, the three mutations that slow desensitization and were used to capture the open state of $hP2X_3$ provide main chain conformational rigidity and make key hydrophobic interactions that stabilize the structure of the cap (Fig. 2c, d). Because these three substitutions are derived from the equivalent wild-type residues in the slowly desensitizing $P2X_2$ receptor, we suggest that the transient formation and stability of the cytoplasmic cap have a central role in P2X receptor gating and provide a structural scaffold for the open state that is likely to be disassembled in the apo and desensitized states.

ATP binding induces cleft closure between the head and dorsal fin domains while pushing the left flipper domain outward[27,43]. These structural rearrangements are transmitted to the lower body, resulting in an outward flexing movement of the β1, β9, β11 and β14 strands. Because the β1 and β14 strands are directly coupled to the TM1 and TM2 helices, respectively (Extended Data Fig. 2a and Fig. 2a, b), their outward flexing pulls on the extracellular portion of the transmembrane domains, causing the helices to expand outward and thereby opening the pore[27,44].

Views from the extracellular side of the membrane, comparing the pore in the apo and open states, show the molecular basis of channel opening (Fig. 2e–g). When the lower body flexes and pulls on TM2, the helix rotates counterclockwise by ~15°. This outward rotation of TM2 promotes the translation of I323, the residue defining the extracellular gate of the apo state, upward by 6.3 Å towards the extracellular surface and reorients it away from the pore centre. The residue that defines the cytoplasmic gate of the apo state, T330, also moves upward by 5.3 Å and rotates away from the pore centre (Fig. 2e–g).

In $zfP2X_4$, the movement of TM2 to open the channel was described as a purely rigid-body transformation[27]. For $hP2X_3$, however, in addition to a rigid-body translation, there is a transition in TM2 from an α-helix to a 3_{10}-helix centred within the sequence G333-V334-G335 (Fig. 2h). The change in helical pitch allows movements of TM2 associated with channel opening and desensitization. We suggest that the formation of the cytoplasmic cap fixes the cytoplasmic portion of TM2 in place, forcing the helix to 'stretch' to a 3_{10} conformation and thus stabilizing pore opening.

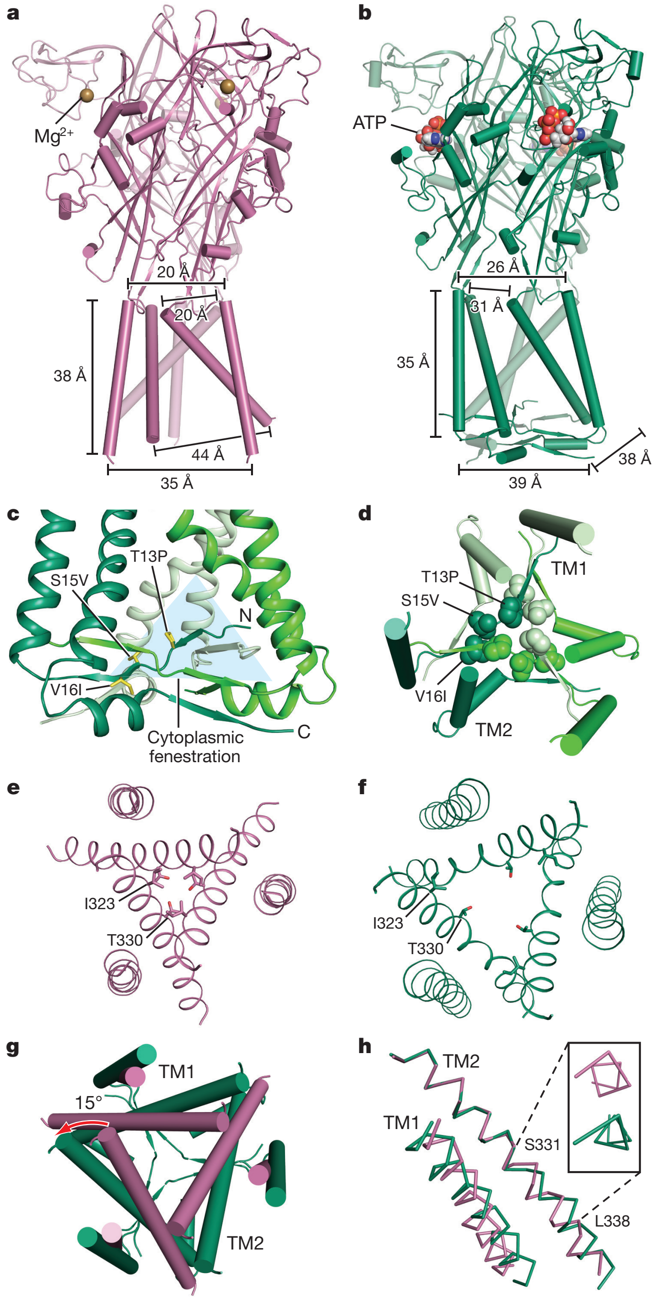


Figure 2 | Apo to open state transition. a, **b**, Apo state (**a**) and open state (**b**) shown parallel to the membrane. The open state structure of $hP2X_3$ visualizes a cytoplasmic motif termed the cytoplasmic cap. **c**, The cytoplasmic cap is composed of domain-swapped β-strands from each protomer, above which are triangular-shaped cytoplasmic fenestrations. Each protomer is coloured in a different shade of green. The T13P, S15V and V16I mutations are shown in one protomer as yellow sticks. **d**, Top-down view from the cytoplasmic surface shows that the residues in the T13P S15V V16I motif form a hydrophobic core. **e**, **f**, Top-down view of the pore comparing the apo state (**e**) to the open state (**f**). **g**, Relative conformational changes in the pore, shown from the extracellular surface, between the apo (red-purple) and open (green) states after aligning the upper body domain of the trimer, demonstrate pore opening. **h**, Alignment of TM2 in apo versus open states reveals a change in helical pitch to a 3_{10}-helix in the open state. The inset shows the view along the axis of the TM2 helix, observed from the cytoplasmic surface.

Channel desensitization

The transmembrane domains and pore architecture differ between the desensitized and open states at the cytoplasmic surface (Figs 2b, 3a). During the transition to the desensitized state, the cytoplasmic portion

of TM2 rotates by about 9° and the short 3_{10}-helix formed in the open state reverts to an α-helix (Fig. 3b, c), resulting in the upward translation towards the extracellular surface (4.4 Å) and inward rotation of V334. This movement in all three protomers closes the pore with V334 redefining a constriction site deeper in the membrane bilayer than the constriction site for the apo state (Fig. 3b, d and Extended Data Fig. 3c).

The transition of TM2 to an ideal helix to close the pore in the desensitized state is not the reverse of the conformational change that opened the pore. The formation of the 3_{10}-helix occurred as a result of stretching of the top half of TM2 upward towards the extracellular surface while its cytoplasmic surface was essentially fixed in place, anchored by the cytoplasmic cap. However, the transition from the open to the desensitized state reverts TM2 to an ideal helix by 'recoiling' the cytoplasmic half of the helix upward. The return of the cytoplasmic half of TM2 to an α-helix resembles the recoiling of a spring. For this recoil movement to occur, the cytoplasmic cap must break or become destabilized to release the 'anchor' and initiate desensitization.

The N terminus in the desensitized state is directed away from the pore, in the opposite direction of the backbone in the open state, suggesting that the structure of cytoplasmic residues differs between these two conformations (Fig. 3e). This finding supports a model in which a transient cytoplasmic cap forms in the open state but ruptures during receptor desensitization. Notably, P2X receptors have a conserved N-terminal glycine[21] (G24 in $P2X_3$) and many subtypes, including $hP2X_3$, have a glycine in the C terminus (G349 in $hP2X_3$) that could act as a hinge[45,46] and provide the flexibility necessary to allow such dynamic conformational changes between functional states

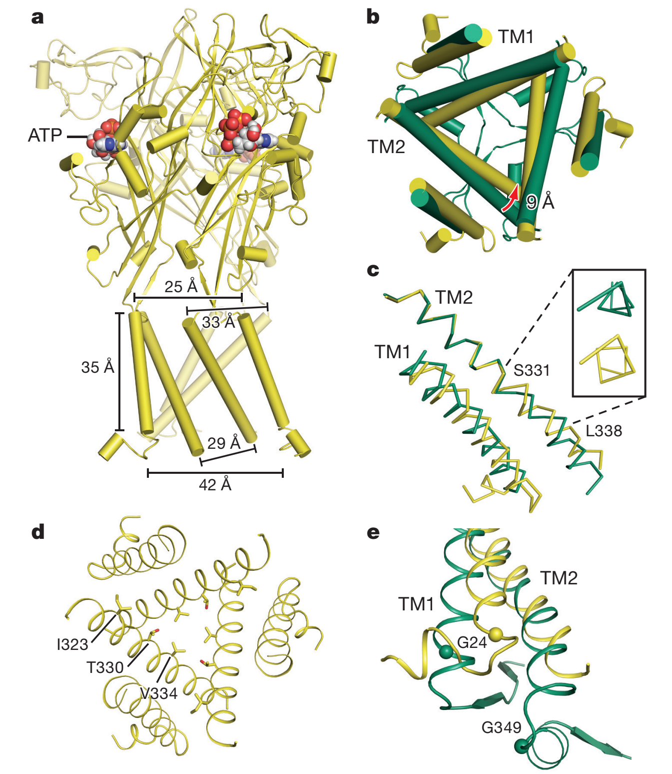


Figure 3 | Open to desensitized state transition. **a**, Structure of the desensitized state shown parallel to the membrane. **b**, Top-down view of the conformational changes in the pore between the open state (green) and the desensitized state (yellow) highlights that the transition to the desensitized state is accompanied by TM2 movement on the cytoplasmic side. **c**, Alignment of TM2 in open versus desensitized states reveals that the 3_{10}-helix in the open state reverts to an α-helix in the desensitized state. The inset shows the view along the axis of the TM2 helix, observed from the cytoplasmic surface. **d**, Top-down view of the pore in the desensitized state. **e**, The C_α atoms of conserved G24 in TM1 of all P2X receptors and G349 in TM2 of $hP2X_3$ are shown as spheres.

(Fig. 3e), as well as the conformational flexibility necessary to 'reset' the receptor to the apo state (Extended Data Fig. 6).

Molecular basis of competitive antagonism

We determined the structure of $hP2X_3$ bound to representatives of two classes of P2X receptor competitive antagonists, TNP-ATP[14,37] and A-317491 (ref. 38). Both ATP and the antagonists occupy the orthosteric ligand-binding pocket, located at the interface between two protomers (Fig. 4 and Extended Data Fig. 7a–f). The most striking difference between ATP and the competitive antagonists is deeper penetration of the latter into the binding cleft. While ATP adopts a U-shape, both TNP-ATP and A-317491 bind in a Y-shape, with the trinitrophenyl moiety of TNP-ATP and the phenoxy-benzyl moiety of A-317491 acting as the 'trunk' (Fig. 4).

In the binding pocket, TNP-ATP adopts a different orientation from ATP. For ATP, the C_2 and C_3 carbons of the ribose group and the γ-phosphate point up, away from the cleft of the binding pocket, whereas for TNP-ATP, these atoms point down, facing into the cleft (Fig. 4d, e and Extended Data Fig. 7g). These differences change how a number of side chain residues interact with the phosphate moieties. For example, K65 makes an ionic interaction with the γ-phosphate of ATP but with the α-phosphate of TNP-ATP. As a result of the different ligand poses, the ribose group of TNP-ATP sits deeper into the cleft of the binding pocket made by the 'left flipper' of protomer A and the 'dorsal fin' of protomer B. The trunk in both TNP-ATP and A-317491 forms hydrophobic interactions with F174, but ATP does not sit deep enough to interact with this residue. By more deeply occupying the space in the cleft between protomers, TNP-ATP and A-317491 prevent the ATP-induced upward movement of the dorsal fin of protomer B to close the binding cleft, precluding the conformational changes necessary for channel opening. TNP-ATP is the prototypical antagonist at $P2X_{1,3}$ receptors but binds substantially less tightly to $P2X_{2,4,7}$ receptors[9,14]. TNP-ATP makes important interactions with K65, D158, T172, F174, N279, R281, and K299 (Fig. 4e). D158 and F174 are not conserved among all P2X family members but both are present in $P2X_{1,3}$, suggesting that the subtype specificity of TNP-ATP is mediated, in part, through these residues.

The apo structure of $hP2X_3$ and both antagonist structures contain a Mg^{2+} ion in the head domain, near the ligand-binding pocket (Extended Data Fig. 7h and Fig. 4e, f), in a different region from that previously predicted[47]. Anomalous difference Fourier maps derived from crystals grown in $MnCl_2$ instead of $MgCl_2$ support the conclusion

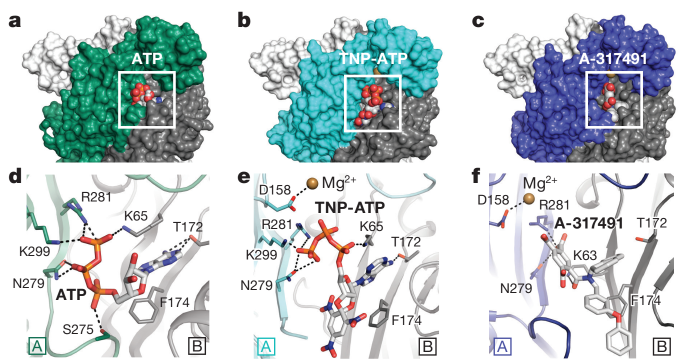


Figure 4 | Orthosteric ligand-binding site. **a**–**c**, Surface representation of the binding pocket for the ATP-bound, open state (**a**), the TNP-ATP-bound, closed state (**b**), and the A-317491-bound, closed state (**c**) of $hP2X_3$. The orthosteric ligands bind in a cleft at an interface between two protomers, with protomer A shown in green for the ATP-bound, open state, cyan for the TNP-ATP-bound, closed state, and blue for the A-317491-bound, closed state. Protomer B is shown in grey and protomer C is shown in white. **d**–**f**, Close-up view of the binding pocket showing key interactions made by ATP (**d**), TNP-ATP (**e**), and A-317491 (**f**). ATP-binding residues make interactions with TNP-ATP and A-317491, notably R281, N279, and K65 and T172 (for TNP-ATP).

that the density feature is a Mg^{2+} ion (Extended Data Fig. 8a, b and Extended Data Table 2). Because of the proximity of Mg^{2+} to the ATP binding site, we investigated whether Mg^{2+} could influence ATP binding affinity, as it has been shown to modulate recovery of receptors from desensitization[48]. However, the presence of Mg^{2+} does not alter the affinity of $hP2X_3$ for ATP (Extended Data Fig. 8c).

Ion access and permeation

On the basis of the $P2X_4$ receptor structure[29], we suggested that ions enter the channel pore from the extracellular milieu via three lateral fenestrations located directly above the transmembrane domains at the extracellular vestibule[49,50] (Extended Data Fig. 3a, b). To investigate how ions enter the channel and identify monovalent cation binding sites, we grew apo $hP2X_3$-MFC_{slow} in the presence of CsCl instead of NaCl and probed for anomalous scattering from Cs^+ ions. An anomalous signal was present in a cavity made by the extracellular vestibule, consistent with Na^+ ions entering through the lateral fenestrations (Fig. 5a, b and Extended Data Table 2).

The egress of ions from the pore of the $hP2X_3$ open state structure to the cytoplasm cannot occur along the threefold axis of the receptor because the orifice along this axis is too small (Extended Data Fig. 3c). However, the cytoplasmic cap and TM2 helices from adjacent protomers form the borders of a triangular-shaped cytoplasmic fenestration, apparently within the boundary of the lipid membrane, that could represent a path of ion egress (Fig. 2c). To test whether these fenestrations are plausible routes for ion egress we carried out molecular dynamic simulations using the open state of the receptor in a 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) lipid bilayer (Fig. 5a, c, d). Hydration patterns in the transmembrane region reveal the putative pathway for ions. Water molecules pass through the open pore but do not exit from the bottom surface of the cytoplasmic cap. Instead, water exits the protein lumen through the cytoplasmic fenestrations (Fig. 5c). Polar lipid head groups line the protein at the fenestrations and probably assist in water permeation. Independent Na^+ permeation events were observed through all three cytoplasmic fenestrations, suggesting that Na^+ ions enter via lateral extracellular fenestrations and egress through lateral cytoplasmic fenestrations (Fig. 5d).

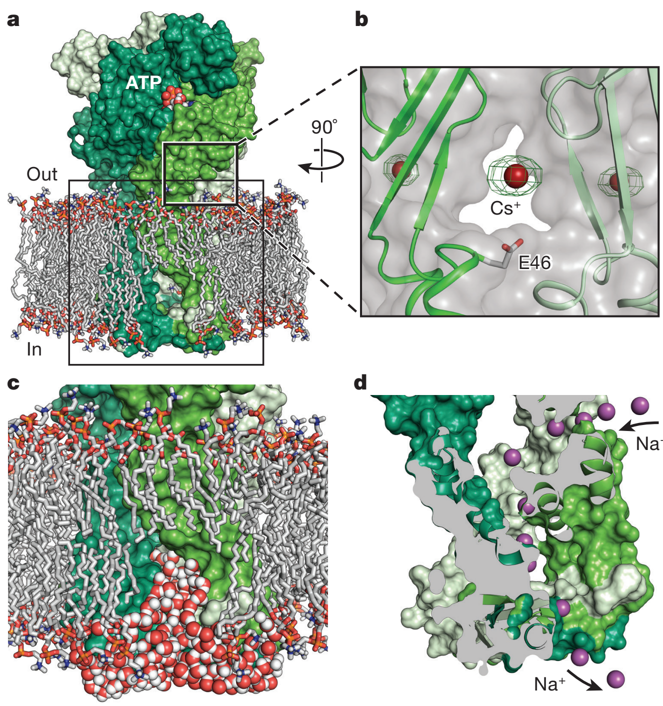


Figure 5 | Extracellular and cytoplasmic fenestrations. a, The equilibrated, membrane-bound model of the open state of $hP2X_3$ with the protein shown in surface representation and each protomer in a different shade of green. POPC lipid tails are silver. For the head groups, oxygen is in red, nitrogen in blue, and phosphorus in orange. **b**, An anomalous peak (5.0σ) for a Cs^+ ion at the entrance of the extracellular vestibule, near E46, which is located at the extracellular end of TM1. This experiment was performed on apo state crystals of $hP2X_3$-MFC_{slow}. **c**, Cytoplasmic fenestrations enable water-filled rivulets, juxtaposed between the protein and lipid membrane, to function as pathways for ion egress into the cytoplasm. Several lipids have been removed in **a** and **c** to allow visualization of the cytoplasmic fenestrations. **d**, Simulation snapshot of an independent Na^+ ion permeation event as Na^+ enters through the extracellular fenestrations and egresses through the cytoplasmic fenestrations. Na^+ ions are shown as purple spheres.

Gating cycle

Initiation of P2X receptor gating begins with the binding of ATP between two subunits, induction of cleft closure and, through structural coupling, an outward flexing of the lower body domain (Supplementary Videos 1–3 and Fig. 6). Because the β-sheets of the lower body domain are directly coupled to the transmembrane helices, their outward movement pulls on the extracellular portions of the transmembrane domains. This conformational change at the extracellular domain induces three major structural changes in the transmembrane and cytoplasmic domains during the transition from the apo to the open state: a counterclockwise rotation of TM2 to open the pore in an iris-like movement; a change in helical pitch for a turn of TM2 from an α-helix to a 3_{10}-helix; and formation of the cytoplasmic cap, which anchors the cytoplasmic surfaces of the transmembrane domains, and provides cytoplasmic fenestrations through which ions exit the pore.

Transition from the open to the desensitized state has two major features: the cytoplasmic cap unfolds or disassembles, and TM2 recoils upward, reverting the short stretch of 3_{10}-helix to an α-helix and allowing the pore to close at a new constriction site, located deeper within the membrane bilayer. In this way, the transition of TM2 from the open state to the desensitized state resembles the recoiling of a spring that has been stretched from above and subsequently released from below. We refer to this as the 'helical recoil' model of receptor desensitization and suggest that the structure of the cytoplasmic cap stabilizes the open state, with its stability tuning the rate and extent of receptor desensitization. The role of the cytoplasmic cap in receptor function is not surprising because residues in both termini of P2X receptors have long been implicated in modulating desensitization[22,23,25].

Conclusion

The structures of all iconic functional states of $hP2X_3$ receptor highlight how the ion pathway changes from the apo to open to desensitized states. We visualize the full-length transmembrane domains and characterize the structural role of the intracellular residues in P2X receptor gating. Our structures reveal how the cytoplasmic cap anchors the transmembrane domain to allow a change in helical pitch in TM2 upon channel opening and provides a phospholipid-lined pathway for ions to laterally exit the pore. We hypothesize that the cytoplasmic cap undergoes a folding–unfolding transition during channel gating and its stability sets the rate of receptor desensitization, with the fast-desensitizing P2X receptor subtypes having a relatively less stable cap domain and the slow and incompletely desensitizing receptor subtypes having a more stable cap domain. The competitive antagonists

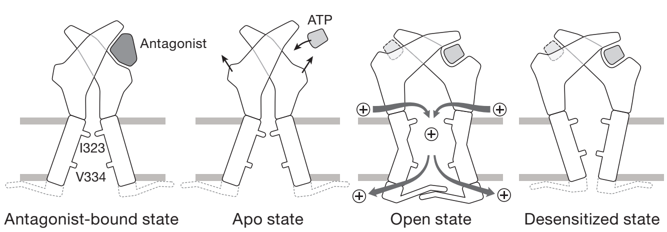


Figure 6 | The gating cycle. A cartoon model summarizing the mechanisms of activation, desensitization, ion permeation/egress and antagonist action of P2X receptors.

TNP-ATP and A-317491 bind to the orthosteric site and stabilize the apo state of the receptor. The structures of $hP2X_3$ represent each of the major conformational states in the receptor gating cycle and illuminate the molecular mechanisms behind P2X receptor activation, desensitization and inhibition.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 4 May; accepted 12 August 2016.
Published online 14 September 2016.

1. Londos, C., Cooper, D. M. & Wolff, J. Subclasses of external adenosine receptors. *Proc. Natl Acad. Sci. USA* **77,** 2551–2554 (1980).
2. van Calker, D., Müller, M. & Hamprecht, B. Adenosine regulates via two different types of receptors, the accumulation of cyclic AMP in cultured brain cells. *J. Neurochem.* **33,** 999–1005 (1979).
3. Burnstock, G. Do some nerve cells release more than one transmitter? *Neuroscience* **1,** 239–248 (1976).
4. Valera, S. *et al.* A new class of ligand-gated ion channel defined by P2x receptor for extracellular ATP. *Nature* **371,** 516–519 (1994).
5. Webb, T. E. *et al.* Cloning and functional expression of a brain G-protein-coupled ATP receptor. *FEBS Lett.* **324,** 219–225 (1993).
6. Fountain, S. J. & Burnstock, G. An evolutionary history of P2X receptors. *Purinergic Signal*. **5,** 269–272 (2009).
7. Burnstock, G. & Kennedy, C. P2X receptors in health and disease. *Adv. Pharmacol.* **61,** 333–372 (2011).
8. Surprenant, A. & North, R. A. Signaling at purinergic P2X receptors. *Annu. Rev. Physiol.* **71,** 333–359 (2009).
9. North, R. A. & Jarvis, M. F. P2X receptors as drug targets. *Mol. Pharmacol.* **83,** 759–769 (2013).
10. Brake, A. J., Wagenbach, M. J. & Julius, D. New structural motif for ligand-gated ion channels defined by an ionotropic ATP receptor. *Nature* **371,** 519–523 (1994).
11. North, R. A. Molecular physiology of P2X receptors. *Physiol. Rev.* **82,** 1013–1067 (2002).
12. Habermacher, C., Dunning, K., Chataigneau, T. & Grutter, T. Molecular structure and function of P2X receptors. *Neuropharmacology* **104,** 18–30 (2016).
13. Egan, T. M. & Khakh, B. S. Contribution of calcium ions to P2X channel responses. *J. Neurosci.* **24,** 3413–3420 (2004).
14. Virginio, C., Robertson, G., Surprenant, A. & North, R. A. Trinitrophenyl-substituted nucleotides are potent antagonists selective for P2X1, P2X3, and heteromeric P2X2/3 receptors. *Mol. Pharmacol.* **53,** 969–973 (1998).
15. Jarvis, M. F. & Khakh, B. S. ATP-gated P2X cation-channels. *Neuropharmacology* **56,** 208–215 (2009).
16. Koshimizu, T., Koshimizu, M. & Stojilkovic, S. S. Contributions of the C-terminal domain to the control of P2X receptor desensitization. *J. Biol. Chem.* **274,** 37651–37657 (1999).
17. Allsopp, R. C. & Evans, R. J. The intracellular amino terminus plays a dominant role in desensitization of ATP-gated P2X receptor ion channels. *J. Biol. Chem.* **286,** 44691–44701 (2011).
18. Allsopp, R. C., Farmer, L. K., Fryatt, A. G. & Evans, R. J. P2X receptor chimeras highlight roles of the amino terminus to partial agonist efficacy, the carboxyl terminus to recovery from desensitization, and independent regulation of channel transitions. *J. Biol. Chem.* **288,** 21412–21421 (2013).
19. Boué-Grabot, E., Archambault, V. & Séguéla, P. A protein kinase C site highly conserved in P2X subunits controls the desensitization kinetics of P2X(2) ATP-gated channels. *J. Biol. Chem.* **275,** 10190–10195 (2000).
20. Brändle, U. *et al.* Desensitization of the P2X(2) receptor controlled by alternative splicing. *FEBS Lett.* **404,** 294–298 (1997).
21. Hausmann, R. *et al.* A hydrophobic residue in position 15 of the rP2X3 receptor slows desensitization and reveals properties beneficial for pharmacological analysis and high-throughput screening. *Neuropharmacology* **79,** 603–615 (2014).
22. Koshimizu, T., Tomić, M., Koshimizu, M. & Stojilkovic, S. S. Identification of amino acid residues contributing to desensitization of the P2X2 receptor channel. *J. Biol. Chem.* **273,** 12853–12857 (1998).
23. Smith, F. M., Humphrey, P. P. & Murrell-Lagnado, R. D. Identification of amino acids within the P2X2 receptor C-terminus that regulate desensitization. *J. Physiol. (Lond.)* **520,** 91–99 (1999).
24. Werner, P., Seward, E. P., Buell, G. N. & North, R. A. Domains of P2X receptors involved in desensitization. *Proc. Natl Acad. Sci. USA* **93,** 15485–15490 (1996).
25. Zemkova, H., He, M. L., Koshimizu, T. A. & Stojilkovic, S. S. Identification of ectodomain regions contributing to gating, deactivation, and resensitization of purinergic P2X receptors. *J. Neurosci.* **24,** 6968–6978 (2004).
26. Fryatt, A. G. & Evans, R. J. Kinetics of conformational changes revealed by voltage-clamp fluorometry give insight to desensitization at ATP-gated human P2X1 receptors. *Mol. Pharmacol.* **86,** 707–715 (2014).
27. Hattori, M. & Gouaux, E. Molecular mechanism of ATP binding and ion channel activation in P2X receptors. *Nature* **485,** 207–212 (2012).
28. Kasuya, G. *et al.* Structural insights into divalent cation modulations of ATP-gated P2X receptor channels. *Cell Reports* **14,** 932–944 (2016).
29. Kawate, T., Michel, J. C., Birdsong, W. T. & Gouaux, E. Crystal structure of the ATP-gated P2X(4) ion channel in the closed state. *Nature* **460,** 592–598 (2009).
30. Habermacher, C. *et al.* Photo-switchable tweezers illuminate pore-opening motions of an ATP-gated P2X ion channel. *eLife* **5,** e11050 (2016).
31. Heymann, G. *et al.* Inter- and intrasubunit interactions between transmembrane helices in the open state of P2X receptor channels. *Proc. Natl Acad. Sci. USA* **110,** E4045–E4054 (2013).
32. Zhou, H. X. & Cross, T. A. Influences of membrane mimetic environments on membrane protein structures. *Annu. Rev. Biophys.* **42,** 361–392 (2013).
33. Minato, Y. *et al.* Conductance of P2X4 purinergic receptor is determined by conformational equilibrium in the transmembrane region. *Proc. Natl Acad. Sci. USA* **113,** 4741–4746 (2016).
34. Quick, M. & Javitch, J. A. Monitoring the function of membrane transport proteins in detergent-solubilized form. *Proc. Natl Acad. Sci. USA* **104,** 3603–3608 (2007).
35. Chen, C. C. *et al.* A P2X purinoceptor expressed by a subset of sensory neurons. *Nature* **377,** 428–431 (1995).
36. Lewis, C. *et al.* Coexpression of P2X2 and P2X3 receptor subunits can account for ATP-gated currents in sensory neurons. *Nature* **377,** 432–435 (1995).
37. Burgard, E. C. *et al.* Competitive antagonism of recombinant P2X(2/3) receptors by 2′, 3′-O-(2,4,6-trinitrophenyl) adenosine 5′-triphosphate (TNP-ATP). *Mol. Pharmacol.* **58,** 1502–1510 (2000).
38. Jarvis, M. F. *et al.* A-317491, a novel potent and selective non-nucleotide antagonist of P2X3 and P2X2/3 receptors, reduces chronic inflammatory and neuropathic pain in the rat. *Proc. Natl Acad. Sci. USA* **99,** 17179–17184 (2002).
39. Egan, T. M., Haines, W. R. & Voigt, M. M. A domain contributing to the ion channel of ATP-gated P2X2 receptors identified by the substituted cysteine accessibility method. *J. Neurosci.* **18,** 2350–2359 (1998).
40. Hille, B. *Ion Channels of Excitable Membranes 3rd edn (Sinauer, 2001).*
41. Degrève, L., Vechi, S. M. & Junior, C. Q. The hydration structure of the Na^+ and K^+ ions and the selectivity of their ionic channels. *Biochim. Biophys. Acta* **1274,** 149–156 (1996).
42. Migita, K., Haines, W. R., Voigt, M. M. & Egan, T. M. Polar residues of the second transmembrane domain influence cation permeability of the ATP-gated P2X(2) receptor.*J. Biol. Chem.* **276,** 30934–30941 (2001).
43. Jiang, R. *et al.* Tightening of the ATP-binding sites induces the opening of P2X receptor channels. *EMBO J.* **31,** 2134–2143 (2012).
44. Li, M., Kawate, T., Silberberg, S. D. & Swartz, K. J. Pore-opening mechanism in trimeric P2X receptor channels. *Nat. Commun.* **1,** 44 (2010).
45. Fujiwara, Y., Keceli, B., Nakajo, K. & Kubo, Y. Voltage- and [ATP]-dependent gating of the P2X(2) ATP receptor channel. *J. Gen. Physiol.* **133,** 93–109 (2009).
46. Khakh, B. S., Bao, X. R., Labarca, C. & Lester, H. A. Neuronal P2X transmitter-gated cation channels change their ion selectivity in seconds. *Nat. Neurosci.* **2,** 322–330 (1999).
47. Li, M., Silberberg, S. D. & Swartz, K. J. Subtype-specific control of P2X receptor channel signaling by ATP and Mg^{2+}. *Proc. Natl Acad. Sci. USA* **110,** E3455–E3463 (2013).
48. Giniatullin, R., Sokolova, E. & Nistri, A. Modulation of P2X3 receptors by Mg^{2+} on rat DRG neurons in culture. *Neuropharmacology* **44,** 132–140 (2003).
49. Kawate, T., Robertson, J. L., Li, M., Silberberg, S. D. & Swartz, K. J. Ion access pathway to the transmembrane pore in P2X receptor channels. *J. Gen. Physiol.* **137,** 579–590 (2011).
50. Samways, D. S., Khakh, B. S., Dutertre, S. & Egan, T. M. Preferential use of unobstructed lateral portals as the access route to the pore of human ATP-gated ion channels (P2X receptors). *Proc. Natl Acad. Sci. USA* **108,** 13800–13805 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank M. Hattori for initial construct screening, L. Vaskalis for figures, H. Owen for manuscript preparation and Gouaux laboratory members for discussions. We acknowledge the Berkeley Center for Structural Biology at the Advanced Light source for assistance with data collection at beamline 5.0.2 and the Northeastern Collaborative Access Team at the Advanced Photon Source for assistance with data collection at beamline 24-ID-C. The simulations were supported by the National Institutes of General Medical Sciences (U54-GM087519 and P41-GM104601 to E.T.) and computationally through XSEDE (TG-MCA06N060 to E.T.). E.G. is an investigator with the Howard Hughes Medical Institute. This research was supported by the National Institute of General Medical Sciences (5F32GM108391 to S.E.M and R01GM100400 to E.G.).

Author Contributions S.E.M. and E.G. designed the project. S.E.M. performed the biochemical and functional analyses. S.E.M. and W.O. carried out the protein purification and crystallization. S.E.M., W.L., and W.O. performed the crystallography and model building. M.S. and E.T. performed the molecular dynamics simulations. All authors wrote and edited the manuscript.

Author Information The coordinates for the structure have been deposited in the Protein Data Bank under the accession codes 5SVJ, 5SVK, 5SVL, 5SVM, 5SVP, 5SVQ, 5SVR, 5SVS, and 5SVT. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to E.G. (gouauxe@ohsu.edu).

Reviewer Information *Nature* thanks P. Biggin, R. Murrell-Lagnado and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Receptor constructs. The initial construct for the $hP2X_3$ receptor was engineered based on the crystallization construct for the open state structure of $zfP2X_4$ ($\Delta P2X_4$-C)[27] and had 19 residues removed from the N terminus and 49 residues removed from the C terminus ($hP2X_3$-ΔN19ΔC49). Although this receptor construct bound ATP with nanomolar affinity in radio-ligand binding assays, it showed no functional gating properties, as assessed by two-electrode voltage clamp experiments. Therefore, we systematically added residues back to both the N and C termini to obtain a functional $hP2X_3$ construct. The return of 14 residues to the N terminus and 16 residues to the C terminus yielded $hP2X_3$-ΔN5ΔC33, referred to as $hP2X_3$-MFC (minimal functional construct). To increase the likelihood of obtaining an open state conformation of the receptor, three rat $P2X_2$-specific amino acid substitutions (P19, V21 and I22) were made at the corresponding positions in the N terminus of human $P2X_3$ to confer the slowly desensitizing receptor phenotype[21], referred to as hP2X3-ΔN5ΔC33-T13P/S15V/V16I or $hP2X_3$-MFC_{slow}.

Expression, membrane preparation and protein purification. The $hP2X_3$-MFC and $hP2X_3$-MFC_{slow} proteins were expressed in HEK293S *GNTI*$^-$ (*GNTI* is also known as *MGAT1*) cells as N-terminal EGFP fusions with an octa-histidine affinity tag and a thrombin cleavage sequence using baculovirus-mediated gene transduction of mammalian cells[51]. HEK293S *GNTI*$^-$ cells in suspension were grown to a density of $3.0 \times 10^6\,ml^{-1}$ and then infected by P2 BacMam virus. After growth at 37 °C for 16 h, sodium butyrate was added to 10 mM final concentration and the cells were shifted to 30 °C for an additional 72 h. Cells were then harvested, washed with PBS buffer and resuspended in TBS (50 mM Tris, pH 8.0 and 150 mM NaCl). The cells were broken by sonication in the presence of protease inhibitors (1 mM PMSF, 0.05 mg/ml aprotinin, 2 μg/ml pepstatin A, and 2 μg/ml leupeptin) and the membrane fraction was isolated by ultracentrifugation.

Pelleted membranes were resuspended in TBS buffer + 15% glycerol, homogenized and solubilized in 40 mM dodecyl-β-D-maltopyranoside, referred to as C12M. The solubilized fraction was isolated by ultra-centrifugation and incubated with TALON resin at 4 °C for 1–2 h. After the resin was packed into an XK-16 column, the column was washed with 12 column volumes of buffer (TBS buffer plus 1 mM C12M, 30 mM imidazole and 5% glycerol) before being eluted with buffer containing 250 mM imidazole, pH 8.0. Fractions were pooled together and the pH was lowered to 6.5 by the addition of 500 mM MES, pH 6.5. Protein was then digested with thrombin (1:25, w/w) and Endo H (1:3, w/w) at room temperature for ~16 h. The digested protein was concentrated and clarified by ultracentrifugation. The supernatant was injected onto a Superdex 200 10/300 GL column pre-equilibrated with 20 mM HEPES, pH 7.0, 100 mM NaCl, and 0.5 mM C12M to isolate trimeric receptors using size-exclusion chromatography (SEC). Monodispersed fractions were collected and $hP2X_3$ was concentrated to 2–3 mg/ml before crystallization. For crystallization experiments designed to locate Na^+ binding sites using the anomalous signal from Cs^+ ions, SEC was performed with 100 mM CsCl instead of 100 mM NaCl.

For apo state and antagonist-bound structures, membranes were subjected to a dialysis step before purification in order to ensure removal of endogenous ATP. To do this, cell membranes were homogenized and transferred into an 8–10 kDa molecular mass cut-off cellulose ester dialysis tubing and dialysed in 150× volume of buffer containing 50 mM Tris pH 9.5, 1 M NaCl, 5% glycerol with buffer exchanges occurring once or twice per day over the course of six days. Dialysed membranes were then solubilized and the protein was purified, as described above.

Crystallization and structure determination. All crystals were obtained with protein at 2–3 mg/ml and set up at 4 °C in the hanging drop vapour diffusion method by mixing 1:1 (v/v) ratio with reservoir buffer. Crystals typically grew after 2–3 weeks.

Crystallization of apo state. Initial experiments to crystallize $hP2X_3$ in the apo state did not include a high salt dialysis step during purification. Structures obtained without dialysis contained a strong density in the binding pocket, consistent with the size, shape, and orientation of ATP, and thus were in fact not apo states of the receptor. We attributed the observed density to endogenous, cellular ATP, which presumably bound during cell lysis and stayed bound throughout the purification. Introducing a high salt dialysis step (see above) allowed the removal of endogenous ATP and the structure determination of a true apo state of the receptor. Purified $hP2X_3$-MFC_{slow} obtained from dialysed membranes was set up with reservoir buffer containing 25% PEG 400, 100 mM MES, pH 6.85, and 50 mM $MgCl_2$. Crystals were cryo-protected by increasing the concentration of PEG 400 to 36% before freezing in liquid nitrogen. For experiments designed to locate putative Mg^{2+} binding sites, crystallization of the apo receptor was performed with 50 mM $MnCl_2$ substituted in place of 50 mM $MgCl_2$.

TNP-ATP soaking experiments. TNP-ATP was added to the drop of apo $hP2X_3$-MFC_{slow} receptor crystals to a final concentration of 1–2 mM and allowed to soak for 24 h before harvesting. Cryo-protection was performed by increasing the concentration of PEG 400 to 36% before freezing in liquid nitrogen.

Crystallization of A-317491-bound state. Purified $hP2X_3$-MFC_{slow} obtained from dialysed membranes was supplemented with 3–4 mM A-317491 and set up in reservoir buffer containing 20% PEG 400, 100 mM glycine, pH 8.5 and 150 mM $MgCl_2$. Crystals were cryo-protected by increasing the concentration of PEG 400 to 36% before freezing in liquid nitrogen.

Crystallization of ATP-bound closed pore state. Purified $hP2X_3$-MFC was supplemented with 0.25 mM TNP-ATP and set up in reservoir buffer containing 21% PEG 400, 100 mM Tris, pH 8.0, 325 mM sodium acetate, and 100 mM NaCl. Crystals were cryo-protected by transfer to well buffer supplemented with 25% ethylene glycol before freezing in liquid nitrogen. Under these conditions, TNP-ATP acted as an additive to facilitate crystallization of $hP2X_3$ bound to endogenous ATP. Exposing these crystals to 1 mM TNP-ATP resulted in their destruction but exposing them to 1 mM ATP or 2-methylthio-ATP kept them intact. No crystals could be grown using TNP-ATP co-crystallization with apo protein where the endogenous ATP had been removed by extensive dialysis (see above).

2-methylthio-ATP soaking experiments. Crystals of $hP2X_3$-MFC grown under the ATP-bound, closed pore conditions were subjected to soaking with either 2-methylthio ATP or TNP-ATP. Crystals were harvested from their drops individually using loops and transferred to a 5-μl drop of reservoir solution plus either 0.5 mM TNP-ATP or 0.5 mM 2-methylthio-ATP and 0.5 mM C12M. The crystals were allowed to soak in the drop for 48 h and crystals that survived were transferred to a second 5-μl drop of reservoir solution with 0.5 mM soaking ligand and 0.5 mM C12M. The crystals were then harvested after 96 h by transfer into reservoir buffer with 0.5 mM ligand, 0.5 mM C12M and 25% ethylene glycol as the cryo-protectant.

Crystallization of ATP-bound open pore state. Purified $hP2X_3$-MFC_{slow} was supplemented with 0.25 mM TNP-ATP and set up in reservoir buffer containing 20% PEG 400 and 50 mM ADA, pH 6.5. Crystals were cryo-protected by increasing the concentration of PEG 400 to 36% before freezing in liquid nitrogen.

Structure determination. X-ray data sets were collected at the Advanced Light Source (beam line 5.0.2) and at the Advanced Photon Source (beam line 24-ID-C). Images were integrated with XDS and scaled with XSCALE[52]. For both the TNP-ATP soaked structure and the A-317491-bound structure, diffraction data were further processed by micro-diffraction data assembly analysis[53]. For the apo state structure, anisotropic scaling was performed with $F/\sigma F = 3.0$ as the cut-off criterion using the UCLA anisotropy server[54]. All structures were solved by molecular replacement using the PHASER package[55,56]. The first $hP2X_3$ structure solved in this study used the apo state $zfP2X_4$ structure as the search model[29]. All subsequent structures, however, used $hP2X_3$ models as search models. Models were built and refined using tools in the CCP4[57], COOT[58], and PHENIX[59] packages. Stereochemistry was evaluated using MolProbity[60]. The three anomalous diffraction data sets were collected at wavelengths near the X-ray absorption edges of the f'' of the element (Extended Data Table 2). The Calculate Maps utility of PHENIX was used to calculate the anomalous maps with high-resolution cut-offs of 4.0 Å, 4.0 Å, and 3.8 Å for the sulfur, manganese and caesium data, respectively.

Two-electrode voltage clamp. RNAs encoding $hP2X_3$-MFC and $hP2X_3$-MFC_{slow} were transcribed from pCDNA3.1x plasmids using the mMessage mMachine T7 Ultra kit. *Xenopus* oocytes were then injected with 5–10 ng RNA and incubated at 18 °C for 1–2 days in a solution containing 96 mM NaCl, 2 mM KCl, 1 mM $MgCl_2$, 1.8 mM $CaCl_2$, 5 mM HEPES, pH 7.5 and 250 μg/ml amikacin. Current recordings were made in a buffer containing 90 mM NaCl, 1 mM KCl, 2 mM $MgCl_2$, and 10 mM HEPES, pH 7.4. Recording electrode pipettes (1–2.5 MΩ) were filled with 3 M KCl. Oocytes were voltage clamped at −60 mV. Traces were recorded with ATP at 1 μM concentration with or without co-application of antagonist, either TNP-ATP or A-317491, at 2 μM. Analogue data were filtered at 50 Hz and digitized at >1 kHz. The Axoclamp 2B amplifier and pClamp 10 software were used for data acquisition.

Radioligand-binding experiments. Scintillation proximity assays (SPA) were performed on detergent-solubilized $hP2X_3$-MFC and $hP2X_3$-MFC_{slow} receptors[34], purified without tag cleavage from dialysed membranes. ATP affinity experiments were carried out using polyvinyltoluene copper (PVT-Cu) beads at 0.5 mg/ml, ^3H-labelled ATP (1:4 ratio, hot ^3H-ATP: cold ATP), and 10 nM GFP-His8-$hP2X_3$ protein in PBS buffer, pH 8.0, 0.3% BSA, and 0.2 mM C12M. The background, non-specific counts were determined by measuring the SPA signal in the presence of 10 μM TNP-ATP. Experiments to test the effect of magnesium ions on ATP affinity were performed in a sulfate buffer (10 mM Tris, pH 8.0, 137 mM NaCl, 2.7 mM KCl, 10 mM sodium sulfate, 1.8 mM potassium sulfate) with or without 25 mM $MgCl_2$.

K_i inhibition studies were carried out using polyvinyltoluene copper (PVT-Cu) beads at 0.5 mg/ml, 10 nM total ATP (2 nM ^3H-labelled ATP: 8 nM cold ATP), and

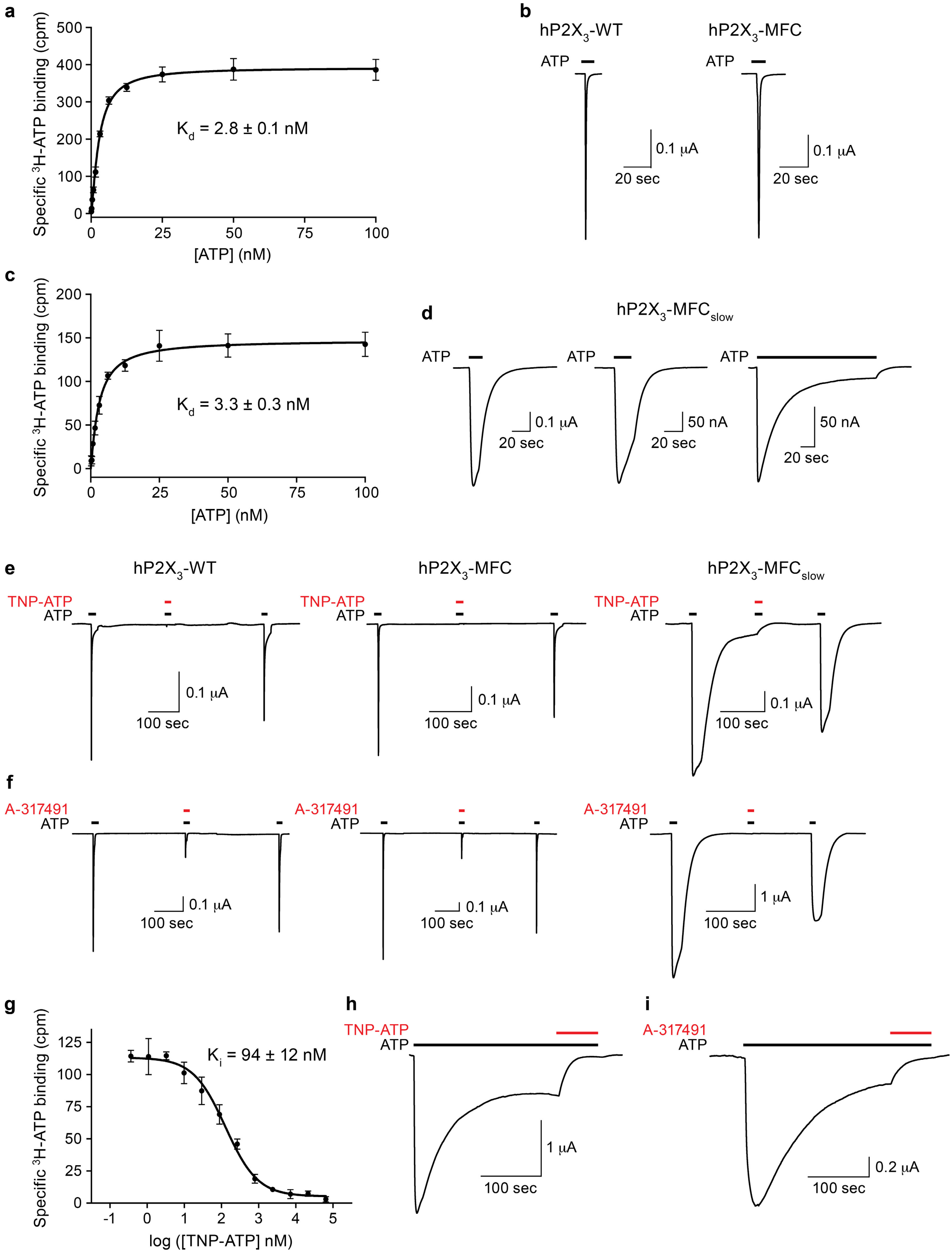
10 nM GFP-His8-hP2X_3 protein in PBS buffer, pH 8.0, 0.3% BSA, and 0.15 mM C12M. Counts were recorded with increasing concentration of cold competitive antagonist. The SPA signal was counted at various time points after gentle agitation (15 min) at room temperature. Assay plates were read using a MicroBeta counter. Data were fitted using a standard single site competition equation, and K_i values were calculated from the IC_{50} values using the Cheng-Prusoff equation. All data points are from triplicate experiments.

Molecular dynamics simulations. *Simulation setup.* The open state of hP2X_3 was used as the initial structure for the simulations. Protonation states of the titratable residues were assigned based on pK_a calculations using PROPKA 3.1[61–63] at pH 7. Accordingly, all the glutamate and aspartate residues were modelled in their default (unprotonated) form. Protein was placed into a POPC lipid bilayer using the replacement method in CHARMM-GUI and solvated[64]. Na^+ and Cl^- ions were added to neutralize the system with a net concentration of 100 mM. The system (197,531 atoms) was minimized for 5,000 steps and simulated for 10 ns at 310 K with all heavy atoms of the protein restrained to their crystallographic positions with a spring constant of $k = 5$ kcal/mol/Å^2. Thereafter, the restraints on the side chains were removed and the system was simulated for an additional 10 ns with only the C_α atoms restrained at a spring constant of $k = 5$ kcal/mol/Å^2. Finally, all the restraints were removed and the system was simulated for 500 ns. The final system had dimensions of 110 Å × 110 Å × 153 Å.

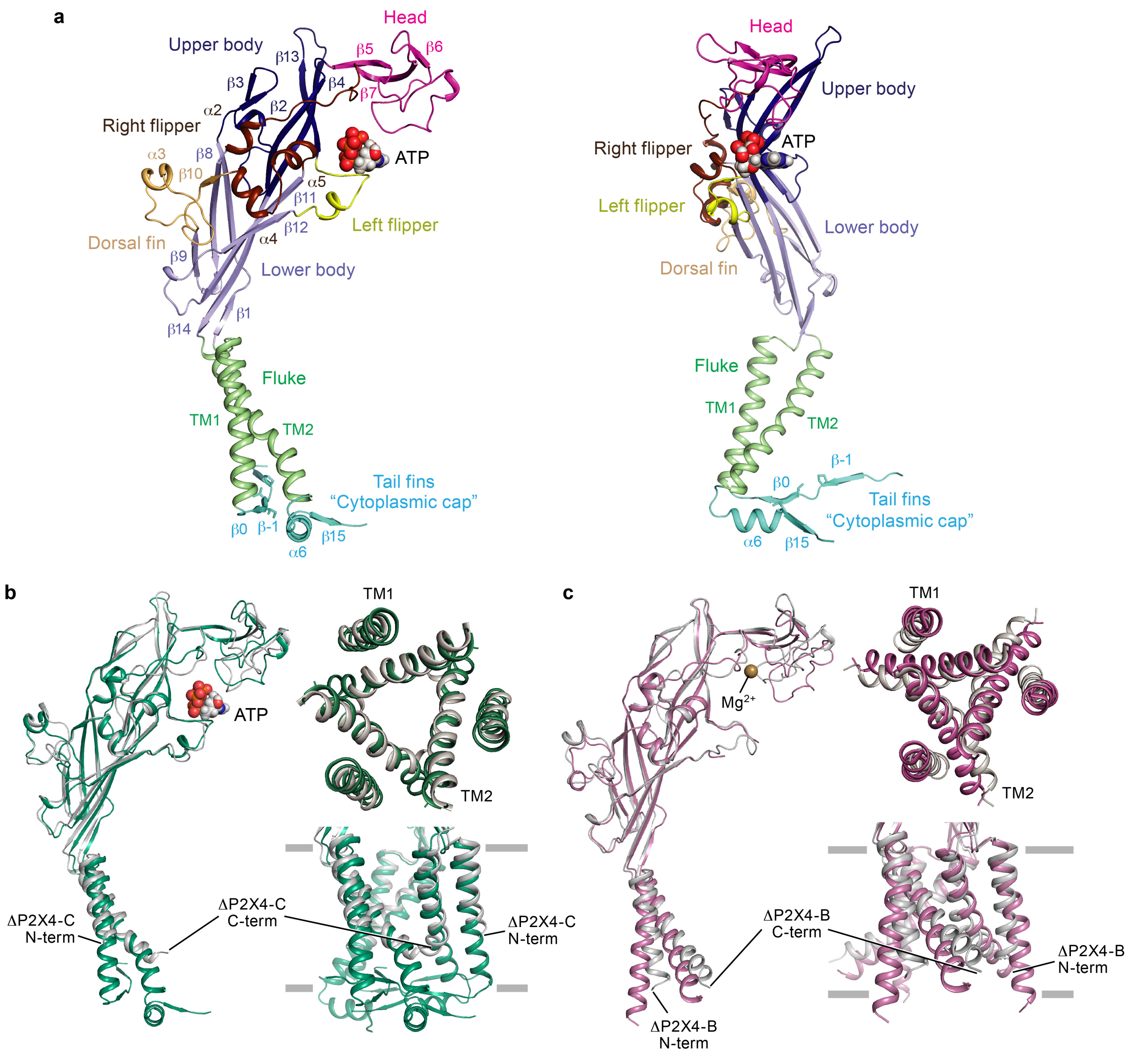
Simulation protocol. All the simulations were performed with NAMD 2.9[65] using CHARMM27 force field for proteins with φ/ψ cross term map (CMAP) corrections[66,67] and CHARMM36 all-atom additive parameters for lipids[68]. Water was modelled as TIP3P[69]. All simulations were performed using the NPT ensemble with periodic boundary conditions. Temperature was maintained at 310 K using Langevin dynamics[70] with a damping coefficient of 0.5 ps^{-1}. Pressure was kept at 1 atm using the Nosé-Hoover Langevin piston method[70,71] with a piston period of 100 fs and a piston decay of 50 fs. Short-range interactions were cut off at 12 Å with a switching applied at 10 Å. Long-range electrostatic forces were calculated using the particle mesh Ewald (PME)[72] method at a grid density of >1 Å^{-3}. Bonded, non-bonded, and PME calculations were performed at 2-, 2-, and 4-fs intervals, respectively. All restraints were in harmonic form with a spring constant of $k = 5$ kcal/mol/Å^2. Minimizations employed a conjugate gradient algorithm.

Simulation under an electric potential. In order to achieve more efficient sampling of the hydrated pathways identified at the protein–lipid interface during the equilibrium simulations, and to assess their potential role as an ion translocation pathway, an independent simulation was set up in which a constant electric potential was applied across the lipid bilayer by imposing a uniform electric field **E** on all atoms of the system along the membrane normal (z-axis). The imposed electric field resulted in a membrane potential difference of 1 V (calculated as $E \cdot l_z$, where E is the magnitude of the electric field and l_z is the length of the periodic cell along the membrane normal). The starting point for the membrane potential simulation was a snapshot at $t = 200$ ns from the equilibrium simulation, which was then simulated under electric potential for an additional 200 ns.

51. Goehring, A. *et al.* Screening and large-scale expression of membrane proteins in mammalian cells for structural studies. *Nat. Protocols* **9,** 2574–2585 (2014).
52. Kabsch, W. Xds. *Acta Crystallogr. D Biol. Crystallogr*. **66,** 125–132 (2010).
53. Hanson, M. A. *et al.* Crystal structure of a lipid G protein-coupled receptor. *Science* **335,** 851–855 (2012).
54. Strong, M. *et al.* Toward the structural genomics of complexes: crystal structure of a PE/PPE protein complex from *Mycobacterium tuberculosis*. *Proc. Natl Acad. Sci. USA* **103,** 8060–8065 (2006).
55. McCoy, A. J. Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallogr. D Biol. Crystallogr*. **63,** 32–41 (2007).
56. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Cryst.* **40,** 658–674 (2007).
57. Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr*. **50,** 760–763 (1994).
58. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr*. **60,** 2126–2132 (2004).
59. Adams, P. D. *et al.* PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D Biol. Crystallogr*. **58,** 1948–1954 (2002).
60. Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr*. **66,** 12–21 (2010).
61. Li, H., Robertson, A. D. & Jensen, J. H. Very fast empirical prediction and rationalization of protein pKa values. *Proteins* **61,** 704–721 (2005).
62. Olsson, M. H., Søndergaard, C. R., Rostkowski, M. & Jensen, J. H. PROPKA3: consistent treatment of internal and surface residues in empirical pKa predictions. *J. Chem. Theory Comput.* **7,** 525–537 (2011).
63. Søndergaard, C. R., Olsson, M. H., Rostkowski, M. & Jensen, J. H. Improved treatment of ligands and coupling effects in empirical calculation and rationalization of pKa values. *J. Chem. Theory Comput.* **7,** 2284–2295 (2011).
64. Jo, S., Kim, T., Iyer, V. G. & Im, W. CHARMM-GUI: a web-based graphical user interface for CHARMM. *J. Comput. Chem.* **29,** 1859–1865 (2008).
65. Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26,** 1781–1802 (2005).
66. MacKerell, A. D. *et al.* All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **102,** 3586–3616 (1998).
67. Mackerell, A. D., Jr, Feig, M. & Brooks, C. L., III. Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* **25,** 1400–1415 (2004).
68. Klauda, J. B. *et al.* Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types. *J. Phys. Chem. B* **114,** 7830–7843 (2010).
69. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79,** 926–935 (1983).
70. Martyna, G. J., Tobias, D. J. & Klein, M. L. Constant pressure molecular dynamics algorithms. *J. Chem. Phys.* **101,** 4177–4189 (1994).
71. Feller, S. E., Zhang, Y., Pastor, R. W. & Brooks, B. R. Constant pressure molecular dynamics simulation: The Langevin piston method. *J. Chem. Phys.* **103,** 4613–4621 (1995).
72. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **98,** 10089–10092 (1993).



Extended Data Figure 1 | Functional studies of hP2X$_3$-MFC and hP2X$_3$-MFC$_{slow}$. **a**, Measurement of [^{3}H-ATP] saturation binding to purified, detergent-solubilized hP2X$_3$-MFC using SPA. For each point in the plot, the error bars indicate the standard error of the mean (SEM) for triplicate samples. The calculated K_d for ATP binding was 2.8 ± 0.1 nM and represents the average of two separate experiments. **b**, ATP-induced currents for hP2X$_3$-WT and hP2X$_3$-MFC both show rapid desensitization kinetics with $\tau = 523 \pm 198$ ms and 429 ± 43 ms, respectively. These values represent an average of three measurements with error values indicating s.e.m. Actual rate constants are likely to be faster as the perfusion rate of our TEVC system is ~1,000 ms. **c**, Measurement of [^{3}H-ATP] saturation binding to purified, detergent solubilized hP2X$_3$-MFC$_{slow}$ using SPA. The calculated K_d for ATP binding was 3.3 ± 0.3 nM. **d**, ATP-induced currents for hP2X$_3$-MFC$_{slow}$ show delayed desensitization kinetics with $\tau = 42{,}581 \pm 2{,}194$ ms. **e**, **f**, Co-application of 2 μM TNP-ATP (**e**) or 2 μM A-317491 (**f**) inhibits the current induced by 1 μM ATP for hP2X$_3$-WT, hP2X$_3$-MFC and hP2X$_3$-MFC$_{slow}$. **g**, Inhibition of ^{3}H-ATP binding to hP2X$_3$-MFC$_{slow}$ by unlabelled TNP-ATP yields a K_i of 94 ± 12 nM. Inhibition of ^{3}H-ATP binding to hP2X$_3$-MFC by unlabelled TNP-ATP yields a K_i of 118 ± 1 nM (data not shown). For each point in the plot, the error bars indicate the s.e.m. for triplicate samples. The reported K_i values represent the average of two separate experiments. **h**, **i**, Co-application of 2 μM TNP-ATP (**h**) or 2 μM A-317491 (**i**) blocks the residual current remaining after prolonged application of 1 μM ATP to hP2X$_3$-MFC$_{slow}$ receptors.



d

Start of hP2X3-MFC

```
hP2X3   (1)    --------MNCISDFFTYETTKSVVVKSWTIGIINRVVQLLIISYFVGWVFLHEKAYQVRD
zfP2X4  (1)    MSESVGCCDSVSQCFFDYYTSKILIIRSKKVGTLNRFTQALVIAYVIGYVCVYNKGYQDTD
```

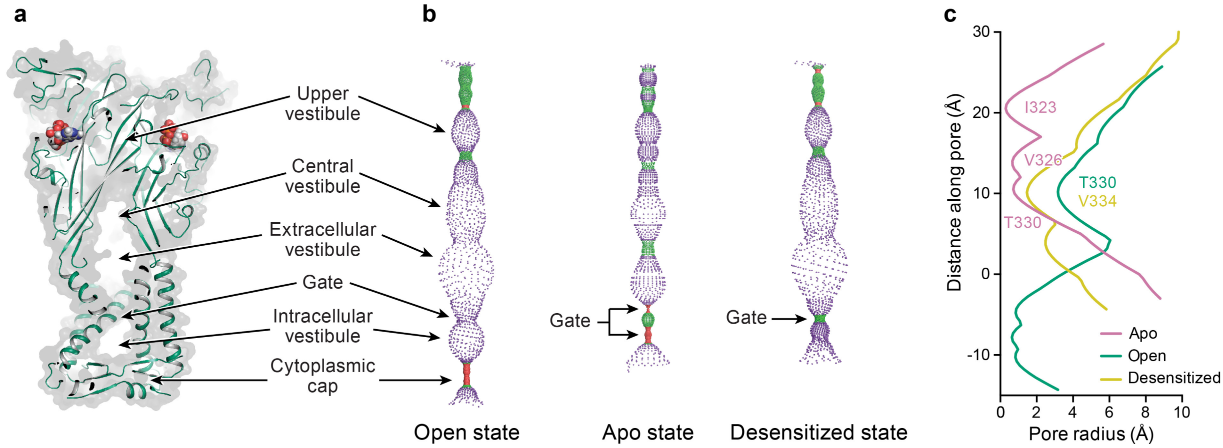
Start of ΔP2X4-C

End of hP2X3-MFC

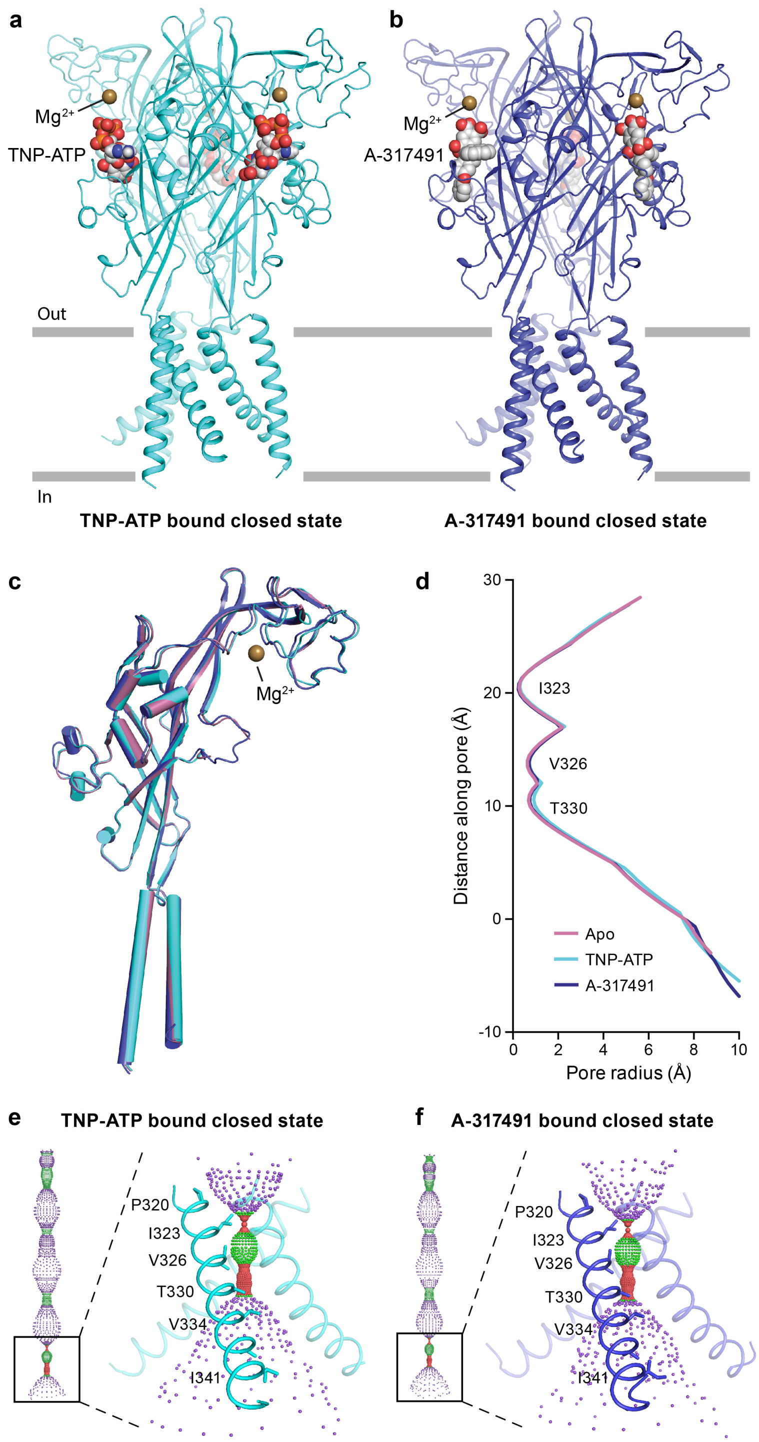
```
hP2X3   (337)  VLCDIILLNFLKGADQYKAKKFEEVNETTLKIAALTNPVYPSDQTTAEKQSTDSGAFSIGH
zfP2X4  (354)  VICDWIVLTFMKRKQHYKEQKYTYVDDFGLLHNEDK-------------------------
```

End of ΔP2X4-C

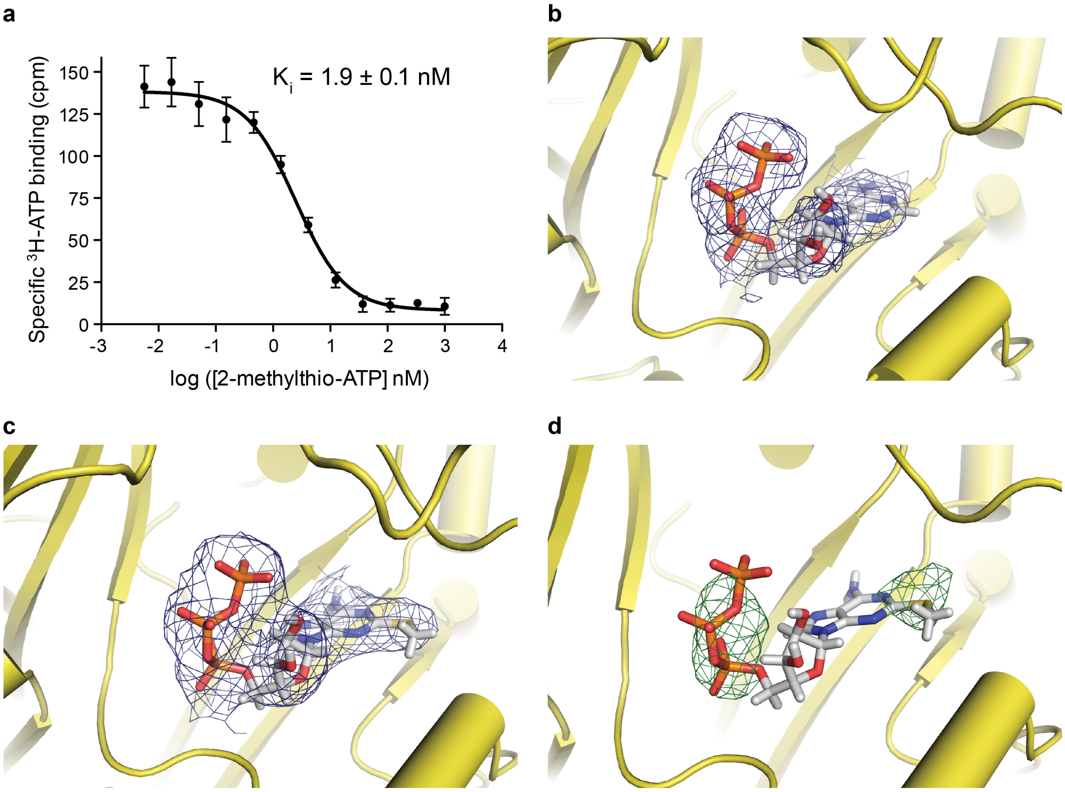
Extended Data Figure 2 | Naming of purinergic receptor domains and comparison of $hP2X_3$ structures to previously published $zfP2X_4$ structures. a, Ribbon representation of one subunit of the open state structure of $hP2X_3$ receptor shown in orthogonal views. The new cytoplasmic cap domain is termed the ‘tail fin’. **b**, Cartoon representation of the open state $hP2X_3$ structure aligned to the open state $zfP2X_4$ structure (construct name ΔP2X4-C) shown parallel to the membrane as a side view and as viewed perpendicular to the membrane from the extracellular side. The transmembrane domains for the $hP2X_3$ structure are longer and more complete than for the $zfP2X_4$ structure. **c**, Cartoon representation of the apo state $hP2X_3$ structure aligned to the apo state $zfP2X_4$ structure (construct name ΔP2X4-B) shown parallel to the membrane as a side view and viewed perpendicular to the membrane from the extracellular side. **d**, Sequence alignment of the N terminus (top alignment) and C terminus (bottom alignment) of $hP2X_3$ compared to $zfP2X_4$. Starting and ending residues of the $hP2X_3$ construct compared to the $\Delta P2X_4$-C construct are indicated with red arrows. The $hP2X_3$ crystallization construct has more residues at both termini than the $\Delta P2X_4$-C crystallization construct.



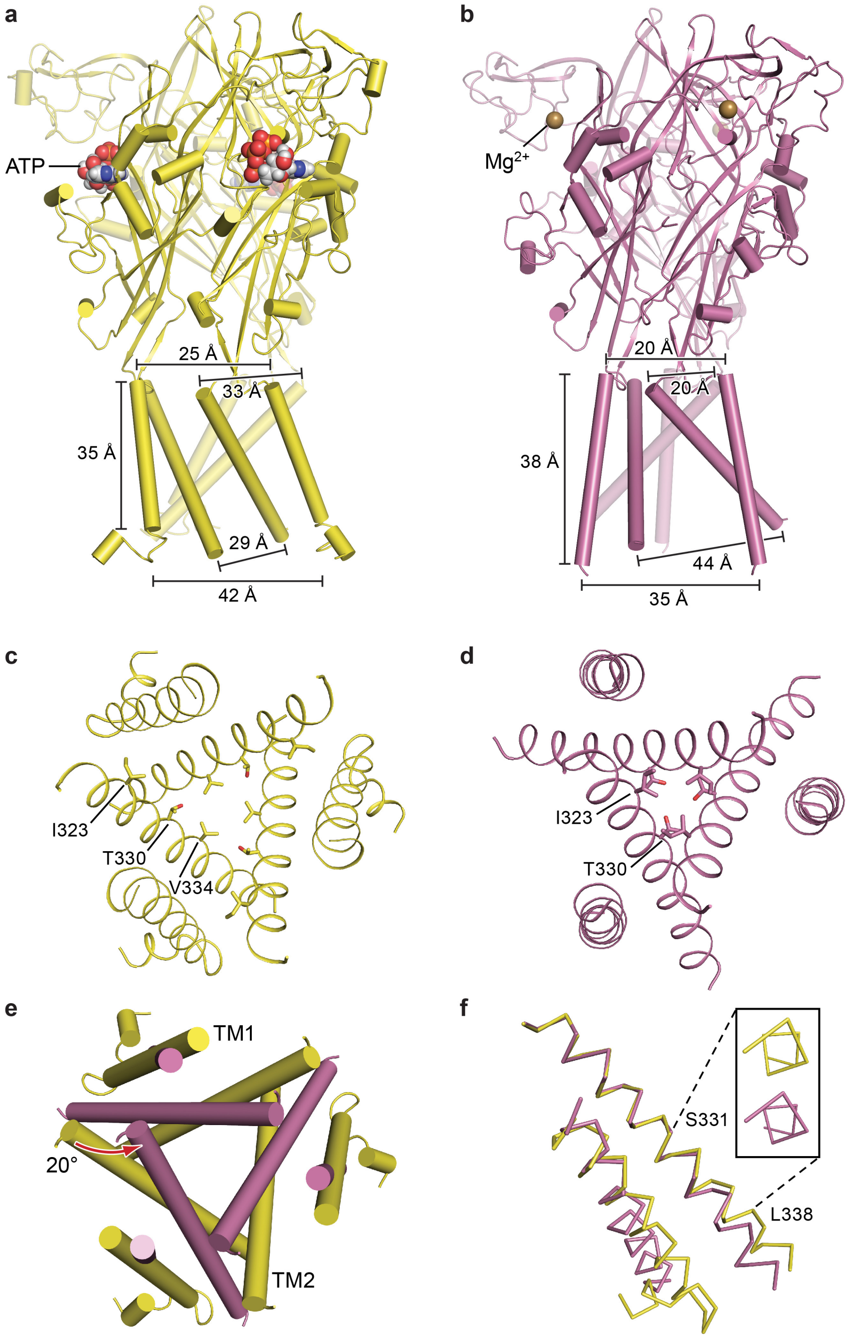
Extended Data Figure 3 | The pore-lining surface of $hP2X_3$ for the open, apo and desensitized states. a, A coronal section of a surface representation of the open state of $hP2X_3$ reveals that four vestibules (upper, central, extracellular and intracellular) are located on the molecular three-fold axis. **b**, Pore-lining surfaces along the entire axis of $hP2X_3$ for open, apo and desensitized states. The colour of each sphere represents a different radius from the receptor centre, as calculated by the program HOLE: red <1.15 Å, green 1.15–2.30 Å, purple >2.30 Å. **c**, Plot of pore radius as a function of distance along the pore axis for the open state versus the apo state versus the desensitized state. The positions of the residues making up the narrowest radius in each conformational state are labelled. The C_α position of I341 is set as zero. I323 defines the first constriction site of the gate (extracellular boundary of the gate), whereas T330 defines the second constriction site (cytoplasmic boundary of the gate). These residues are at the equivalent positions that define the boundaries of the gate in the apo state structure of $zfP2X_4$, but are leucine and alanine residues, respectively, in $zfP2X_4$. A single residue, V334, defines the constriction site of the desensitized state. Residue T330 defines the narrowest region of the pore in the open state.



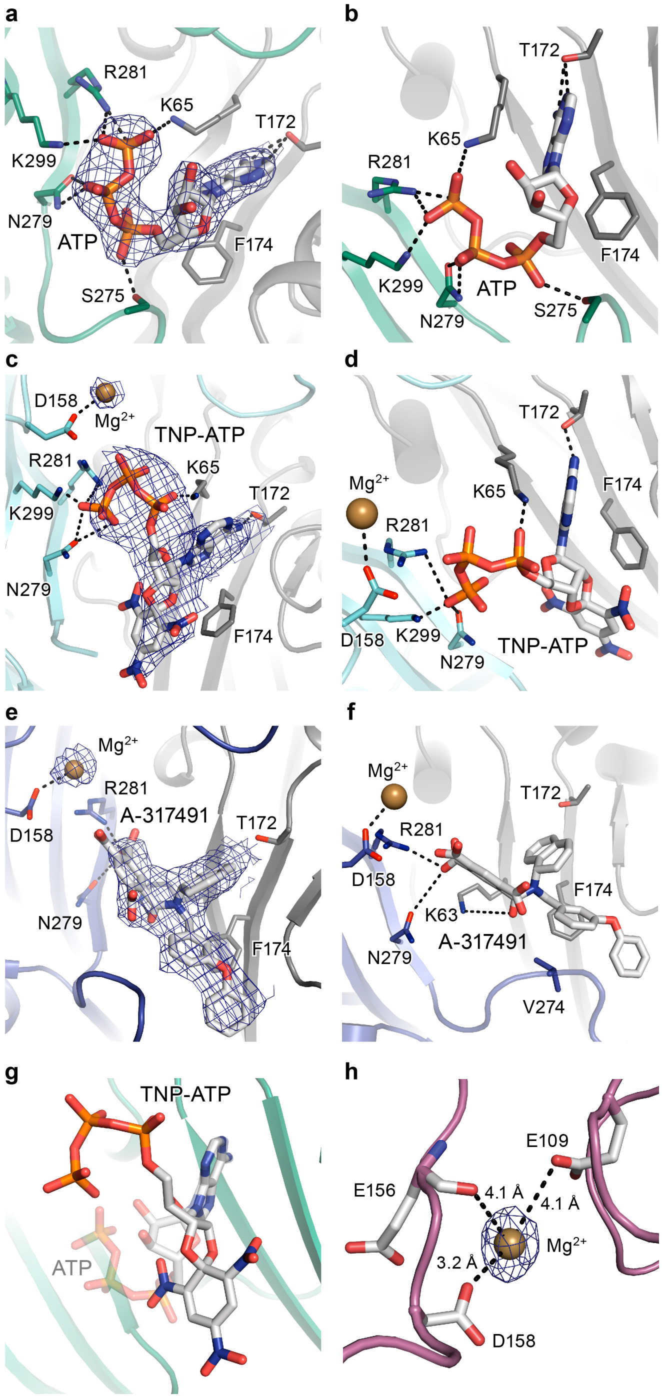
Extended Data Figure 4 | The overall structure and ion channel pore of antagonist-bound/closed states. a, **b**, Cartoon representation of the competitive antagonist-bound/closed state structures, TNP-ATP in cyan (**a**) and A-317491 in blue (**b**), shown parallel to the membrane as a side view. **c**, An overall alignment of a single protomer in the apo state (red-purple), TNP-ATP-bound state (cyan) and A-317491-bound state (blue). **d**, Plot of pore radius as a function of distance along the pore axis for apo state versus TNP-ATP-bound state versus A-317491-bound state. The positions of the residues making up the narrowest radius in each conformational state are labelled. The C_{α} position of I341 is set as zero. **e**, **f**, Pore-lining surfaces along the entire axis of the receptor and a focus on the transmembrane domain with TM2 pore-lining residues shown as sticks for the TNP-ATP-bound state (**e**) and the A-317491-bound state (**f**). The colour of each sphere represents a different radius from the receptor centre, as calculated by the program HOLE: red <1.15 Å, green 1.15–2.30 Å, purple >2.30 Å.



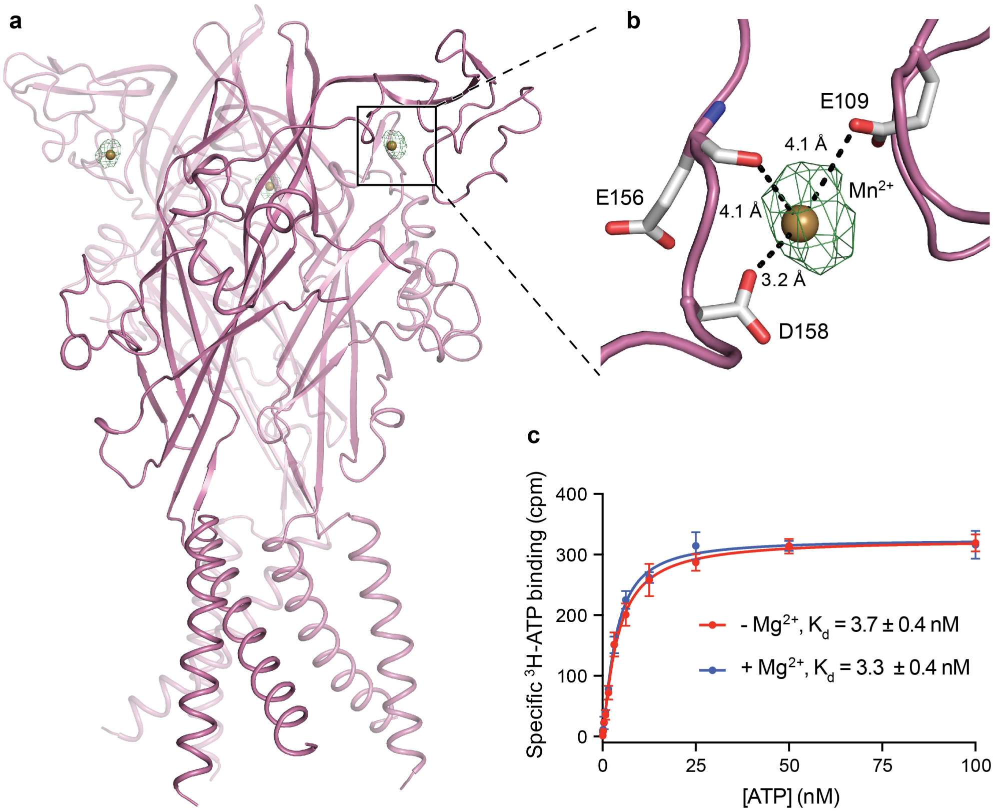
Extended Data Figure 5 | High-affinity P2X$_3$ agonist 2-methylthio-ATP can be soaked into the desensitized state crystals. a, Inhibition of ^{3}H-ATP binding to hP2X$_3$-MFC by unlabelled 2-methylthio-ATP yields a K_i of 1.9 ± 0.1 nM. For each point in the plot, the error bars indicate the s.e.m. for triplicate samples. The reported K_i represents the average of two separate experiments. **b**, Electron density for ATP in the desensitized state. The $F_o - F_c$ map is contoured at 1.0σ. **c**, Desensitized state crystals that have been soaked with 2-methylthio-ATP have a density in the binding pocket, which matches the shape of 2-methylthio-ATP, accounting for the methyl-thio moiety. The $F_o - F_c$ map is contoured at 1.0σ. **d**, An anomalous difference Fourier map (contoured at 3.0σ) has anomalous signal that overlaps with the sulfur moiety of 2-methylthio-ATP as well as the phosphate groups. These crystals of hP2X$_3$-MFC successfully ligand-exchanged ATP for agonist 2-methylthio-ATP in the binding pocket but were destroyed when soaked with antagonist TNP-ATP, providing evidence that the structure represents an agonist-bound, closed or desensitized state.



Extended Data Figure 6 | Resetting from desensitized to apo state of hP2X$_3$. **a**, **b**, Structures of hP2X$_3$ in the desensitized state **(a)** and apo state **(b)** shown parallel to the membrane. There are marked changes between the two states in the extracellular domain and the transmembrane domain. **c**, **d**, Top-down view comparing the pore of the desensitized state **(c)** to the pore of the apo state **(d)** highlighting how, although both pores are closed, the residues that define the gates are different. **e**, Relative differences in the pore between desensitized and apo states after aligning the upper body domain of the trimer reveal that a significant clockwise conformational change at both the extracellular and cytoplasmic surfaces of the transmembrane domain must occur for the receptor pore to reset back to the apo state. **f**, Alignment of TM2 in desensitized versus apo state shows that both helices have the same helical pitch, suggesting that the 3_{10}-helix that existed in the open state is a transient structural feature. The inset shows the view along the axis of the TM2 helix, observed from the cytoplasmic surface. We speculate that the structural resetting of the receptor from the desensitized state to the apo state is likely to occur after ligand dissociation.



Extended Data Figure 7 | Orthosteric ligand-binding site and ligand densities. a, **b**, View of the orthosteric binding pocket for the ATP-bound open state structure of hP2X$_3$. ATP binds at an interface between two protomers, with protomer A shown in green and protomer B shown in grey. The $2F_o-F_c$ density for ATP is shown at 2.5σ. **c**, **d**, View of the orthosteric binding pocket for the TNP-ATP-bound closed state structure of hP2X$_3$ with protomer A shown in cyan and protomer B shown in grey. The $2\,F_o-F_c$ density for TNP-ATP is shown at 1.5σ. **e**, **f**, View of the orthosteric binding pocket for the A-317491-bound closed state structure of hP2X$_3$ with protomer A shown in blue and protomer B shown in grey. The $2F_o-F_c$ density for A-317491 is shown at 0.8σ. **g**, Close-up comparison of the relative orientation of ATP (shown as translucent) versus TNP-ATP in the binding pocket highlights how the phosphate moiety and the orientation of the ribose group are both inverted between the two molecules. **h**, The apo state structure (shown in figure) as well as both antagonist-bound structures have a Mg^{2+} ion present in the head domain of hP2X$_3$, coordinated by the side chains of E109 and D158 as well as the carbonyl oxygen of E156. The $2F_o-F_c$ density for the Mg^{2+} ion is shown at 1.5σ.



Extended Data Figure 8 | Anomalous signal from Mn^{2+} ion proves Mg^{2+} ion is present in the head domain of the apo state. a, Anomalous difference map of apo structure with crystals grown in $MnCl_2$ have an anomalous signal from a Mn^{2+} ion in the head domain (anomalous difference Fourier map shown in green contoured at 5.5σ). This anomalous signal from Mn^{2+} overlaps with the $2F_o-F_c$ density shown in Extended Data Fig. 7h, proving this density is a Mg^{2+} ion. **b**, The Mn^{2+} ion in the head domain is coordinated by the side chains of E109 and D158 and the carbonyl oxygen of E156. **c**, The presence of a Mg^{2+} ion does not change the affinity of ATP for hP2X$_3$-MFC$_{slow}$, as assessed by SPA binding, suggesting that Mg^{2+} does not compete with ATP for the binding pocket or impair the ability of ATP to bind to the receptor. For each point in the plot, the error bars indicate the s.e.m. for triplicate measurements. The reported K_d values represent the mean of two separate experiments.

Extended Data Table 1 | Data collection and refinement statistics

	$hP2X_3$-MFC_{slow} ATP-bound	$hP2X_3$-MFC ATP-bound	$hP2X_3$-MFC_{slow} No Ligand	$hP2X_3$-MFC_{slow} A-317491-bound†	$hP2X_3$-MFC_{slow} No Ligand TNP-ATP Soaked‡	hP2X3-MFC ATP-bound 2-MeThio-ATP Soaked
Data collection	ALS 5.0.2	APS 24-ID-C	ALS 5.0.2	ALS 5.0.2	APS 24-ID-C	APS 24-ID-C
Space group	$P2_13$	$P2_13$	R32	R32	R32	$P2_13$
Cell dimensions						
a, *b*, *c* (Å)	173.15,173.15 173.15	172.14, 172.14, 172.14	120.17, 120.17 236.58	123.17, 123.17, 237.46	120.45, 120.45 235.99	172.64,172.64, 172.64
α, β, γ (°)	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 120.0	90.0, 90.0, 120.0	90.0, 90.0, 120.0	90.0, 90.0, 90.0
Wavelength (Å)	1.0	0.979	1.0	1.0	0.979	0.979
Resolution (Å)	50 – 2.77	80 – 2.90	50 – 2.98	50 – 3.13	50 – 3.25	50 – 3.09
R_{meas}*	8.2 (133.6)	8.1 (120.4)	5.2 (61.5)	6.1 (57.7)	6.1 (59.4)	8.9 (102.1)
$I/\sigma I$*	12.34 (1.22)	12.30 (1.19)	18.34 (2.42)	20.37 (2.91)	19.64 (2.40)	11.10 (1.21)
Completeness (%)*	99.2 (95.9)	97.6 (96.9)	95.4 (58)	98.3 (99.8)	99.8 (99.1)	96.2 (92.8)
Multiplicity*	3.94 (3.87)	2.84 (2.86)	4.12 (5.08)	10.67 (8.96)	10.27 (6.52)	2.32 (2.31)
$CC_{1/2}$ (%)*	99.9 (49.7)	99.8 (43.4)	99.9 (80.8)	100 (32.6)	99.9 (47.6)	99.8 (41.7)
Refinement						
Resolution (Å)*	43 – 2.77 (2.81 – 2.77)	77 – 2.90 (2.95 – 2.90)	39 – 2.98 (3.21 – 2.98)	40 – 3.13 (3.44 – 3.13)	48 – 3.25 (3.58 – 3.25)	48 – 3.09 (3.16 – 3.09)
No. reflections	79845	62472	13067	12308	10651	46746
R_{work}/R_{free}	0.201/0.228	0.205/0.240	0.213/0.259	0.252/0.283	0.254/0.286	0.203/0.228
Average B-factor ($Å^2$)						
Protein	90	89	84	120	125	98
R.m.s. deviations						
Bond lengths (Å)	0.003	0.006	0.003	0.004	0.003	0.003
Bond angles (°)	0.892	1.350	0.772	0.955	0.634	0.721
Ramachandran plot						
Favored (%)	97.6	98.6	97.8	96.6	98.4	98.1
Allowed (%)	2.4	1.4	2.2	3.4	1.6	1.9
Disallowed (%)	0	0	0	0	0	0
Rotamer outliers (%)	0.7	0.7	1.2	0	0.4	0.6

5% of reflections were used for calculation of R_{free}.
*Highest resolution shell in parentheses.
†Two crystals were merged for the A-317491-bound structure and processed with microdiffraction assembly.
‡Two crystals were merged for the TNP-ATP-bound structure and processed with microdiffraction assembly.

Extended Data Table 2 | Anomalous data collection statistics

	$hP2X_3$-MFC ATP-bound 2-MeThio-ATP Soaked Sulfur Anomalous	$hP2X_3$-MFC_{slow} No ligand Mn^{2+} Anomalous	$hP2X_3$-MFC_{slow} No Ligand Cs^{+} Anomalous
Data collection	APS 24-ID-C	ALS 5.0.2	APS 24-ID-C
Space group	$P2_13$	R32	R32
Cell dimensions			
a, *b*, *c* (Å)	172.93, 172.93 172.93	120.20, 120.20 236.24	119.95, 119.95, 236.41
α, β, γ (°)	90.0, 90.0, 90.0	90.0, 90.0, 120.0	90.0, 90.0, 120.0
Wavelength (Å)	1.550	1.771	1.907
Resolution (Å)	50 – 3.30	50 – 4.03	50 – 3.79
R_{meas}*	15.6 (197.8)	14.7 (185.3)	8.2 (292)
$I/\sigma I$*	11.75 (1.33)	11.60 (1.33)	13.42 (0.64)
Completeness (%)*	99.9 (99.0)	99.7 (97.2)	99.8 (99.6)
Multiplicity*	5.82 (5.77)	11.34 (10.14)	7.80 (7.70)
$CC_{1/2}$ (%)*	99.8 (42.6)	100 (57.1)	100 (28.1)

*Highest resolution shell in parentheses.

LETTER

doi:10.1038/nature19320

High-molecular-weight organic matter in the particles of comet 67P/Churyumov–Gerasimenko

Nicolas Fray[1]*, Anaïs Bardyn[1,2]*, Hervé Cottin[1]*, Kathrin Altwegg[3], Donia Baklouti[4], Christelle Briois[2], Luigi Colangeli[5], Cécile Engrand[6], Henning Fischer[7], Albrecht Glasmachers[8], Eberhard Grün[9], Gerhard Haerendel[10], Hartmut Henkel[11], Herwig Höfner[10], Klaus Hornung[12], Elmar K. Jessberger[13], Andreas Koch[11], Harald Krüger[7], Yves Langevin[4], Harry Lehto[14], Kirsi Lehto[15], Léna Le Roy[3], Sihane Merouane[7], Paola Modica[1,2], François-Régis Orthous-Daunay[16], John Paquette[7], François Raulin[1], Jouni Rynö[17], Rita Schulz[18], Johan Silén[17], Sandra Siljeström[19], Wolfgang Steiger[20], Oliver Stenzel[7], Thomas Stephan[21], Laurent Thirkell[2], Roger Thomas[2], Klaus Torkar[22], Kurt Varmuza[23], Karl-Peter Wanczek[24], Boris Zaprudin[14], Jochen Kissel[7] & Martin Hilchenbach[7]

The presence of solid carbonaceous matter in cometary dust was established by the detection of elements such as carbon, hydrogen, oxygen and nitrogen in particles from comet 1P/Halley[1,2]. Such matter is generally thought to have originated in the interstellar medium[3], but it might have formed in the solar nebula—the cloud of gas and dust that was left over after the Sun formed[4]. This solid carbonaceous material cannot be observed from Earth, so it has eluded unambiguous characterization[5]. Many gaseous organic molecules, however, have been observed[6–9]; they come mostly from the sublimation of ices at the surface or in the subsurface of cometary nuclei[8]. These ices could have been formed from material inherited from the interstellar medium that suffered little processing in the solar nebula[10]. Here we report the *in situ* detection of solid organic matter in the dust particles emitted by comet 67P/Churyumov–Gerasimenko; the carbon in this organic material is bound in very large macromolecular compounds, analogous to the insoluble organic matter found in the carbonaceous chondrite meteorites[11,12]. The organic matter in meteorites might have formed in the interstellar medium and/or the solar nebula, but was almost certainly modified in the meteorites' parent bodies[11]. We conclude that the observed cometary carbonaceous solid matter could have the same origin as the meteoritic insoluble organic matter, but suffered less modification before and/or after being incorporated into the comet.

By July 2016, the cometary secondary ion mass analyser (COSIMA) instrument[13] on board the European Space Agency spacecraft Rosetta had detected more than 27,000 particles in the vicinity of comet 67P/Churyumov–Gerasimenko (hereafter 67P). Of them, more than 200 particles have been analysed. They show various morphologies[14,15] and mineral compositions[16], inferred from time-of-flight secondary ion mass spectrometry (TOF-SIMS) analyses. Here we present findings on the organic content of two representative particles, named Kenneth and Juliette. Kenneth was collected between 11 and 12 May 2015, while Juliette was collected between 23 and 29 October 2015. Both were analysed a few weeks after their collection (Extended Data Table 1) and both are larger than 100 μm in size (Extended Data Fig. 1).

Figure 1 shows a comparison of the mass spectra measured on the Kenneth and Juliette particles (in red) and those measured nearby on the porous gold substrates (the 'targets') on which the particles were collected (in black). Mass spectra for five further particles are shown in Extended Data Fig. 2. A comparison of these seven sets of mass spectra—from particles covering the whole range of possible morphologies and collection dates (Extended Data Table 1)—shows that the Kenneth and Juliette particles are representative of the cometary material analysed by COSIMA so far. The positive-ion mass spectra from the cometary particles show signatures of carbon compounds at m/z ratios of 12.00 (C^+), 13.01 (CH^+), 14.02 (CH_2^+) and 15.02 (CH_3^+), with an additional weaker contribution at $m/z = 27.02$ ($C_2H_3^+$). The relative intensities of these ions are different in the spectra measured on the particles compared with those nearby on the same target, showing that most of the C^+ ions are arising from the cometary material (Extended Data Fig. 3). Several previously reported elements[16]—such as sodium, magnesium, aluminium, silicon, calcium and iron—are also observed in the spectra measured on the particles. For some of them, protonated ions can also be detected. Thereby, for example, the peak at mass 29 on Fig. 1 is mainly due to $^{28}SiH^+$. These elements are related to mineral phases. In negative-ion spectra, the dominant carbon signals are at $m/z = 12.00$ (C^-) and 13.01 (CH^-), with much weaker contributions at $m/z = 14.02$ (CH_2^-). Similar results are obtained for all of the cometary particles that have been analysed to date (Fig. 1 and Extended Data Fig. 2). The spectra shown in black display characteristic features of the targets, and are substantially different from the spectra measured on the cometary particles. The ion images in Extended Data Fig. 4 show that the carbon peaks at $m/z = 12.00$, in both positive and negative modes, are all correlated with the particles.

A striking feature of all of the spectra from the 67P particles is that carbon-bearing ions with a cometary origin are observed only at m/z

[1]Laboratoire Interuniversitaire des Systèmes Atmosphériques (LISA), UMR CNRS 7583, Université Paris Est Créteil et Université Paris Diderot, Institut Pierre Simon Laplace, 94000 Créteil, France. [2]Laboratoire de Physique et Chimie de l'Environnement et de l'Espace (LPC2E), CNRS/Université d'Orléans, 45071 Orléans, France. [3]Center for Space and Habitability (CSH), University of Bern, Sidlerstrasse 5, 3012 Bern, Switzerland. [4]Institut d'Astrophysique Spatiale, CNRS/Université Paris Sud, Bâtiment 121, 91405 Orsay, France. [5]European Space Agency, European Space Research and Technology Centre, Keplerlaan 1, Postbus 299, 2200 AG Noordwijk, The Netherlands. [6]Centre de Sciences Nucléaires et de Sciences de la Matière, CNRS/IN2P3 – Université Paris Sud – UMR 8609, Université Paris-Saclay, Bâtiment 104, 91405 Orsay Campus, France. [7]Max-Planck-Institut für Sonnensystemforschung, Justus-von-Liebig-Weg 3, 37077 Göttingen, Germany. [8]Universität Wuppertal, FB-E, Lehrstuhl für Messtechnik, Rainer-Gruenter-Str. 21, 42119 Wuppertal, Germany. [9]Max-Planck-Institut für Kernphysik, Saupfercheckweg 1, 69115 Heidelberg, Germany. [10]Max-Planck-Institut für Extraterrestrische Physik, Giessenbachstrasse, 85748 Garching, Germany. [11]von Hoerner und Sulger GmbH, Schlossplatz 8, 68723 Schwetzingen, Germany. [12]Universität der Bundeswehr LRT-7, Werner Heisenberg Weg 39, 85577 Neubiberg, Germany. [13]Institut für Planetologie, Universität Münster, Berghalde 31f, 69126 Heidelberg, Germany. [14]Tuorla Observatory, Department of Physics and Astronomy, University of Turku, Väisäläntie 20, 21500 Piikkiö, Finland. [15]Laboratory of Molecular Plant Biology, Department of Biochemistry, PharmaCity, Itäinen Pitkäkatu 4B 6 krs, 20520 University of Turku, Finland. [16]Institut de Planétologie et d'Astrophysique de Grenoble, UMR 5274, Université Grenoble Alpes, CNRS, 38000 Grenoble, France. [17]Finnish Meteorological Institute, Observation Services, Erik Palménin Aukio 1, 00560 Helsinki, Finland. [18]European Space Agency, Scientific Support Office, Keplerlaan 1, Postbus 299, 2200 AG Noordwijk, The Netherlands. [19]Department of Chemistry, Materials and Surfaces, SP Technical Research Institute of Sweden, Box 857, 501 15 Borås, Sweden. [20]RC Seibersdorf Research GmbH Business Field Aerospace Technology, 2444 Seibersdorf, Austria. [21]University of Chicago, Department of the Geophysical Sciences, 5734 South Ellis Avenue, Chicago, Illinois 60637, USA. [22]Space Research Institute, Austrian Academy of Sciences, Schmiedlstrasse 6, 8042 Graz, Austria. [23]Institute of Statistics and Mathematical Methods in Economics, Vienna University of Technology, Wiedner Hauptstrasse 7/105-6, 1040 Vienna, Austria. [24]Institut für Anorganische und Physikalische Chemie, FB 2, NW 2, A0110, Universität Bremen, 28334 Bremen, Germany.
*These authors contributed equally to this work.

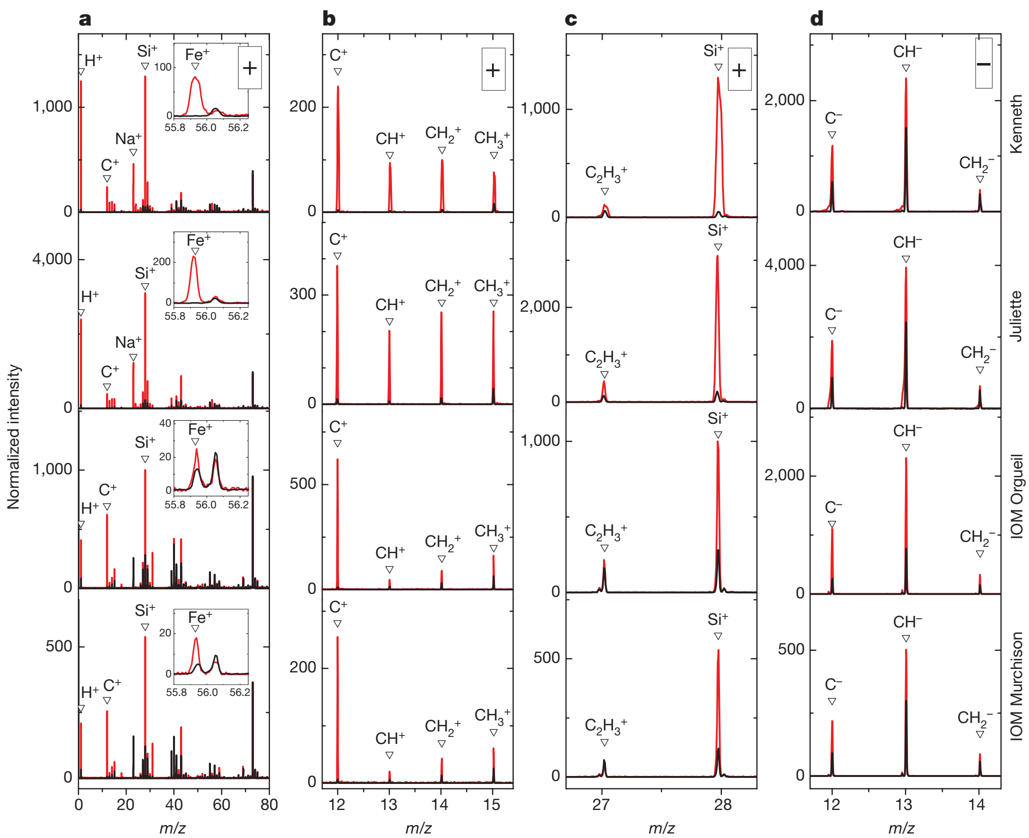


Figure 1 | Comparison of cometary and meteoritic TOF-SIMS data. The spectra in red have been acquired on two cometary particles, Kenneth and Juliette, and on two IOM samples, extracted from the Orgueil and Murchison chondrites. The spectra in black have been acquired on the porous gold substrates on which the cometary particles were collected and IOM samples prepared. The spectra are normalized to the intensity of characteristic fragments of polydimethylsiloxane (PDMS; see Methods). **a**, Positive-ion spectra. **b**, **c**, Enlargements from **a**. **d**, Negative-ion spectra.

ratios of less than 50. In addition, in the positive-ion spectra from the particles, the C^+ peak is the most intense one of the series C^+, CH^+, CH_2^+, CH_3^+ (see Fig. 1 and Extended Data Figs 2, 3). Such spectral characteristics are very different from those observed for the large range of organic molecules with well defined structures that were previously analysed by the COSIMA ground models[17]. These characteristics also indicate a rather low hydrogen/carbon ratio for the detected cometary organic material, in comparison with the ratio for the molecules studied for calibration purposes[17]. This result is in line with the most recent interpretations of spectra from the surface of 67P that were measured by the visible and infrared thermal imaging spectrometer (VIRTIS)[18]. Figure 1 and Extended Data Fig. 3 show that the best analogues found so far to the organic signatures of the 67P particles are the insoluble organic matter (IOM) samples extracted from carbonaceous chondrites (such as the Orgueil and Murchison meteorites). Indeed, as for the spectra of cometary particles, the spectra of those IOMs display a C^+ peak that is the most intense peak of the series C^+, CH^+, CH_2^+, CH_3^+; moreover, no organic ions originating from the IOMs are observed for m/z ratios greater than 50. This resemblance suggests that the cometary carbon in the particles is bound in high-molecular-weight organic matter that bears similarities with the IOM found in the most primitive carbonaceous meteorites[11,12]. However, the mass spectra measured on the cometary particles display some differences compared with those from the IOM samples. In particular, the CH_x^+/C^+ ratios in the spectra are higher for the cometary material than for the IOMs, suggesting a higher hydrogen content in the cometary particles.

Results from laboratory simulations[19] and from analyses of natural analogues such as carbonaceous chondrites[20,21] have indicated that the solid phase of cometary particles should contain a very large variety of organic molecules, ranging from the smallest molecules to high-molecular-weight organic matter. In preparation for the interpretation of mass spectra measured by COSIMA in space, calibration of the COSIMA reference models on Earth was performed on pure organic compounds from a variety of chemical families[17]. The results show that, for about 75% of the investigated organic molecules, the most intense peaks in the positive-ion mass spectra are protonated quasi-molecular ions, $[M+H]^+$. In the negative-ion spectra, the most intense peaks are the fragments C^-, CH^-, O^-, C_2^-, C_2H^-, CN^- and CNO^-, with only some molecules having a detectable deprotonated quasi-molecular ion, $[M-H]^-$, in the mass spectra. Therefore, a large number of peaks was expected in both the positive and the negative spectra of cometary particles, corresponding to various molecules with diverse fragmentation patterns.

While we report evidence of high-molecular-weight organic matter in the particles of 67P, there is no detection by COSIMA of smaller molecules that would be an equivalent to the soluble organic matter found in carbonaceous chondrites (for example, carboxylic acids, aliphatic or polycyclic aromatic hydrocarbons, and amino acids)[17]. Moreover, polyoxymethylene—which has been considered as a possible source of gaseous formaldehyde in cometary atmospheres[22], and tentatively reported by the PTOLEMY instrument at the surface of the nucleus[23]—has not been detected so far in the particles analysed by COSIMA. However, the Rosetta orbital spectrometer for ion and neutral analysis (ROSINA) instrument has detected numerous volatile organic molecules in the gas phase of comet 67P (refs 7, 9). The cometary sampling and composition (COSAC)[24] and PTOLEMY[23] instruments have also reported the detection of volatile organic molecules, which were transferred to the instruments from particles when the Philae lander bounced at the surface of the comet. Those detections are consistent with the first steps of photochemical and/or radiolytic evolution in ices—steps that could have taken place before the comet formed, in the protosolar molecular clouds or on the edge of the solar nebula[9,23,24]. However, there has been no detection of molecules that could be intermediates in chemical evolution, between the previously detected molecules and the high-molecular-weight organic matter reported here. These results suggest different sources for the volatile and refractory carbonaceous material in comet 67P.

In 1986, *in situ* analyses of comet 1P/Halley particles demonstrated their high organic content[1,2], but characterizing this organic matter was a challenge because of the fast flyby velocity (more than $60\,km\,s^{-1}$). Some molecular ions (for example, adenine and toluene) were tentatively reported in comet 1P/Halley[25], but they have not yet

been detected in the COSIMA data. Taking into account the nature of carbon in carbonaceous chondrites and interplanetary dust particles (IDPs), it has been inferred that high-molecular-weight organic matter is probably present in so-called CHON particles (which are composed mainly of the light elements carbon, hydrogen, oxygen and nitrogen)[5]. Our data are compatible with that suggestion. Our data are also compatible with the presence of high-molecular-weight organic matter in IDPs and ultracarbonaceous Antarctic micrometeorites[26,27]. Analysis of cometary particles brought back to Earth by the NASA Stardust space mission[28] showed that a significant fraction of the mineral material studied has been formed in the innermost warm regions of the protoplanetary disk, before being accreted onto the nucleus of comet 81P/Wild 2. However, the organic compounds in these samples were altered by the high-speed collection process[28], hampering comparison with the COSIMA data.

Our results show that the best analogue found so far to the organic refractory component of 67P dust particles is the IOM extracted from carbonaceous chondrites. The similarities between the IOM samples and the cometary spectra could indicate a common origin of the refractory organic matter in these objects. The exact site of this common origin has been extensively discussed, and could be either the interstellar medium[29] or the solar nebula[4,30]. Nevertheless, the hydrogen/carbon ratio of the cometary organic matter in 67P seems to be higher than in the IOMs. Moreover, the hydrogen/carbon ratio in the IOMs of chondrites probably decreases owing to processing of the parent body[11]. As cometary nuclei have not been submitted to such processing since they formed, the higher hydrogen/carbon ratio in cometary particles suggests that the cometary material is more primitive than that in chondritic IOMs.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 6 April; accepted 15 July 2016.
Published online 7 September; corrected online 5 October 2016 (see full-text HTML version for details).

1. Kissel, J. *et al.* Composition of comet Halley dust particles from Giotto observations. *Nature* **321,** 336–337 (1986).
2. Kissel, J. *et al.* Composition of comet Halley dust particles from Vega observations. *Nature* **321,** 280–282 (1986).
3. Greenberg, J. M. Making a comet nucleus. *Astron. Astrophys.* **330,** 375–380 (1998).
4. Nuth, J. A., III, Johnson, N. M. & Manning, S. A self-perpetuating catalyst for the production of complex organic molecules in protostellar nebulae. *Astrophys. J.* **673,** L225–L228 (2008).
5. Fomenkova, M. N., Chang, S. & Mukhin, L. M. Carbonaceous components in the comet Halley dust. *Geochim. Cosmochim. Acta* **58,** 4503–4512 (1994).
6. Biver, N. *et al.* Ethyl alcohol and sugar in comet C/2014 Q2 (Lovejoy). *Sci. Adv.* **1,** e1500863 (2015).
7. Le Roy, L. *et al.* Inventory of the volatiles on comet 67P/Churyumov-Gerasimenko from Rosetta/ROSINA. *Astron. Astrophys.* **583,** A1 (2015).
8. Mumma, M. J. & Charnley, S. B. The chemical composition of comets—emerging taxonomies and natal heritage. *Annu. Rev. Astron. Astrophys.* **49,** 471–524 (2011).
9. Altwegg, K. *et al.* Prebiotic chemicals—amino acid and phosphorus—in the coma of comet 67P/Churyumov-Gerasimenko. *Sci. Adv.* **2,** e1600285 (2016).
10. Bockelée-Morvan, D. *et al.* New molecules found in comet C/1995 O1 (Hale-Bopp). Investigating the link between cometary and interstellar material. *Astron. Astrophys.* **353,** 1101–1114 (2000).
11. Alexander, C. M. O. D., Fogel, M., Yabuta, H. & Cody, G. D. The origin and evolution of chondrites recorded in the elemental and isotopic compositions of their macromolecular organic matter. *Geochim. Cosmochim. Acta* **71,** 4380–4403 (2007).
12. Quirico, E. *et al.* Origin of insoluble organic matter in type 1 and 2 chondrites: new clues, new questions. *Geochim. Cosmochim. Acta* **136,** 80–99 (2014).
13. Kissel, J. *et al.* COSIMA, a high resolution time of flight spectrometer for secondary ion mass spectroscopy of cometary dust particles. *Space Sci. Rev.* **128,** 823–867 (2007).
14. Schulz, R. *et al.* Comet 67P/Churyumov–Gerasimenko sheds dust coat accumulated over the past four years. *Nature* **518,** 216–218 (2015).
15. Langevin, Y. *et al.* Typology of dust particles collected by the COSIMA mass spectrometer in the inner coma of 67P/Churyumov Gerasimenko. *Icarus* **271,** 76–97 (2016).
16. Hilchenbach, M. *et al.* Comet 67P/Churyumov–Gerasimenko: close-up on dust particle fragments. *Astrophys. J.* **816,** L32 (2016).
17. Le Roy, L. *et al.* COSIMA calibration for the detection and characterisation of the cometary solid organic matter. *Planet. Space Sci.* **105,** 1–25 (2015).
18. Quirico, E. *et al.* Refractory and semi-volatile organics at the surface of comet 67P/Churyumov-Gerasimenko: insights from the VIRTIS/Rosetta imaging spectrometer. *Icarus* **272,** 32–47 (2016).
19. Danger, G. *et al.* Characterization of laboratory analogs of interstellar/cometary organic residues using very high resolution mass spectrometry. *Geochim. Cosmochim. Acta* **118,** 184–201 (2013).
20. Schmitt-Kopplin, P. *et al.* High molecular diversity of extraterrestrial organic matter in Murchison meteorite revealed 40 years after its fall. *Proc. Natl Acad. Sci. USA* **107,** 2763–2768 (2010).
21. Sephton, M. A. Organic compounds in carbonaceous meteorites. *Nat. Prod. Rep.* **19,** 292–311 (2002).
22. Cottin, H. & Fray, N. Distributed sources in comets. *Space Sci. Rev.* **138,** 179–197 (2008).
23. Wright, I. P. *et al.* CHO-bearing organic compounds at the surface of 67P/Churyumov-Gerasimenko revealed by Ptolemy. *Science* **349,** aab0673 (2015).
24. Goesmann, F. *et al.* Organic compounds on comet 67P/Churyumov-Gerasimenko revealed by COSAC mass spectrometry. *Science* **349,** aab0689 (2015).
25. Kissel, J. & Krueger, F. R. The organic component in dust from comet Halley as measured by the PUMA mass spectrometer on board Vega 1. *Nature* **326,** 755–760 (1987).
26. Flynn, G. J., Keller, L. P., Feser, M., Wirick, S. & Jacobsen, C. The origin of organic matter in the solar system: evidence from the interplanetary dust particles. *Geochim. Cosmochim. Acta* **67,** 4791–4806 (2003).
27. Dartois, E. *et al.* Ultracarbonaceous Antarctic micrometeorites, probing the Solar System beyond the nitrogen snow-line. *Icarus* **224,** 243–252 (2013).
28. Brownlee, D. The Stardust mission: analyzing samples from the edge of the Solar System. *Annu. Rev. Earth Planet. Sci.* **42,** 179–205 (2014).
29. Alexander, C. M. O. D., Cody, G. D., Fogel, M., Yabuta, H. & Sanford, S. in *Organic Matter in Space* Vol. 251 (ed. Kwok, S.) 293–298 (2008).
30. Ciesla, F. J. & Sandford, S. A. Organic synthesis via irradiation and warming of ice grains in the Solar nebula. *Science* **336,** 452–454 (2012).

Acknowledgements COSIMA was built by a consortium led by the Max-Planck-Institut für Extraterrestrische Physik, Garching, Germany, in collaboration with: the Laboratoire de Physique et Chimie de l'Environnement et de l'Espace, Orléans, France; the Institut d'Astrophysique Spatiale, CNRS/Université Paris Sud, Orsay, France; the Finnish Meteorological Institute, Helsinki, Finland; the Universität Wuppertal, Wuppertal, Germany; von Hoerner und Sulger GmbH, Schwetzingen, Germany; the Universität der Bundeswehr, Neubiberg, Germany; the Institut für Physik, Forschungszentrum Seibersdorf, Seibersdorf, Austria; and the Institut für Weltraumforschung, Österreichische Akademie der Wissenschaften, Graz, Austria; and is led by the Max-Planck-Institut für Sonnensystemforschung, Göttingen, Germany. We acknowledge the support of the national funding agencies of Germany (Deutsches Zentrum für Luft- und Raumfahrt (DLR), grant 50 QP 1302), France (Centre National d'Étude Spatiales (CNES)), Austria, Finland and the European Space Agency (ESA) Technical Directorate. A.B. acknowledges support from the CNES and the Labex Exploration Spatiale des Environnements Planétaires (ESEP; no. 2011-LABX-030), and funding from the Idex Paris Sciences et Lettres (PSL; no. ANR-10-IDEX-0001-02). S.S. acknowledges support from the Swedish National Space Board grant (contracts 121/11 and 198/15). H.L., B.Z. and K.L. acknowledge Academy of Finland grant 277375. We thank the Rosetta Science Ground Segment at the European Space Astronomy Centre, the Rosetta Mission Operations Centre at the European Space Operations Centre, and the Rosetta Project at the European Space Research and Technology Centre for their work, which enabled the scientific return of the Rosetta mission.

Author Contributions N.F. and A.B. carried out the first identification of carbon in the particles presented here, and contributed to interpreting the results. N.F. provided the figures. H.C. contributed to the interpretation, and wrote the manuscript with the assistance of A.B. and N.F. J.K. and M.H. are the two successive managers of the project. K.A., L.C., A.G., E.G., G.H., H.He., H.Hö., K.H., E.K.J., J.K., A.K., Y.L., F.R., J.R., R.S., J.S., W.S., T.S., L.T., R.T., K.T., K.V. and K.-P.W. contributed to instrument development. H.F., M.H., A.K., Y.L., J.P., J.R., O.S. and L.T. contributed to instrument operations and data distribution. D.B., A.B., C.B., H.C., C.E., N.F., M.H., K.H., H.K., J.K., H.L., K.L. Y.L., L.L.R., S.M., P.M., F.-R.O.-D., J.P., J.R., J.S., S.S., O.S., L.T. and B.Z. contributed to instrument calibration and data analysis. Y.L. provided grain pictures. All authors discussed the results and commented on the manuscript.

Author Information After the proprietary period of six months, the data will be available in the ESA Planetary Science Archive (see Methods). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to N.F. (nicolas.fray@lisa.u-pec.fr).

Reviewer Information *Nature* thanks C. Alexander and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

The COSIMA instrument was designed to capture cometary dust particles at low velocity (a few meters per second) in order to preserve their molecular integrity. This can be compared to a capture velocity of $6\,km\,s^{-1}$ for grains captured by the Stardust space mission[31,32]; a velocity impact of $78\,km\,s^{-1}$ for the PUMA instruments on board the Vega spacecrafts[25]; and a velocity impact of $68\,km\,s^{-1}$ for the PIA instrument on board the Giotto spacecraft[1].

Cometary dust particles are collected by COSIMA on sets of three targets of various materials, each $10\,mm \times 10\,mm$ in size[33], exposed simultaneously outside the spacecraft. The dust particles discussed here were collected on porous gold substrates at temperatures of about 283 K. After imagery, selected particles are probed by TOF-SIMS, a technique that analyses the top surface layers of solid samples. The mass resolution, $m/\Delta m$, of COSIMA is about 1,400 (full-width half-maximum at $m/z = 100$), which allows, in most cases, elements from hydrogen-containing organic ions of the same integer mass to be separated, thanks to mass defect properties. The indium-ion beam is pulsed as required for TOF-SIMS; its footprint is $35 \times 50\,\mu m^2$ (full-width at half-maximum). Each pulse lasts less than 3 ns and contains about 1,000 $^{115}In^+$ ions with an energy of 8 keV (ref. 13).

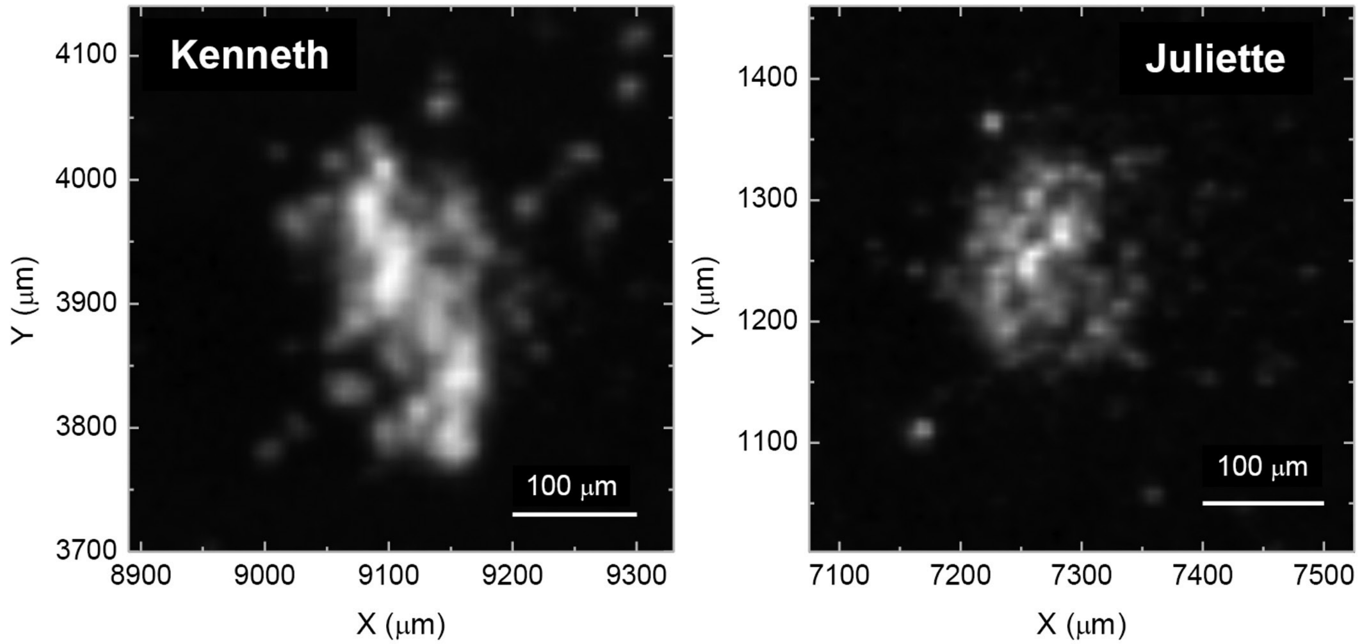
Depending on the sampled location, individual mass spectra contain secondary ions from the cometary particle, from the target surface, or from both. Positive- and negative-ion spectra are normalized to the intensities of the peaks at $m/z = 73.05$ ($Si(CH_3)_3{}^+$) and at $m/z = 59.00$ ($SiOCH_3{}^-$), respectively; these are characteristic fragments of polydimethylsiloxane (PDMS)—an organic molecule, originating from the instrument's background, that is extremely efficiently ionized. In addition to gold and indium, PDMS fragments are characteristic features of the target and therefore a non-cometary component of each spectrum. Normalization to PDMS allows spectra to be compared. The large increase in the carbon/PDMS ratios in the positive-ions spectra acquired on the particles compared with those acquired nearby on the same target (Fig. 1 and Extended Data Figs 2– 4), as well as the change in the relative intensities of C^+, CH^+, $CH_2{}^+$ and $CH_3{}^+$ ions, indicates that the identified signatures of carbon on the particles have a cometary origin. To increase the signal-to-noise ratio of the carbon signal, we added several spectra. The spectra shown in Fig. 1 and Extended Data Figs 2 and 3 correspond to the addition of several spectra (between 2 and 100) that were measured in different locations on the same particles.

In this study, we analysed samples a few weeks after they were collected. The particle Kenneth was collected between 11 and 12 May 2015 and was later analysed twice: on 18 and 18 June 2015 and on 1 and 2 July 2015. The particle Juliette was collected between 23 and 29 October 2015 and was analysed on 18 November 2015 (see Extended Data Table 1). Thus, any possible volatile fraction (including ices) present at the time of particle collection had plausibly been lost by the time of the first analyses.

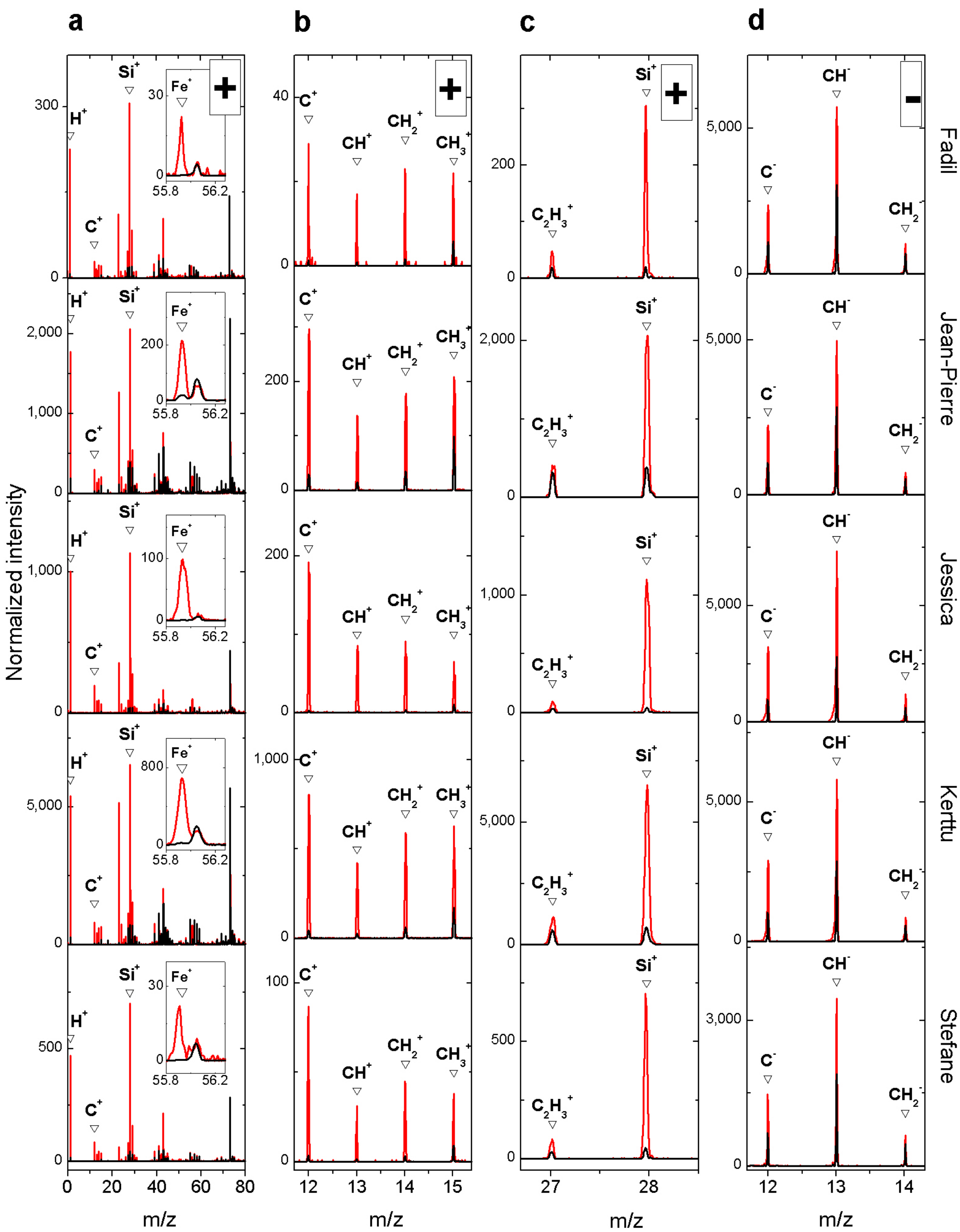
All COSIMA data have been, or will be, released to the Planetary Science Archive of ESA (http://www.cosmos.esa.int/web/psa/psa-introduction) and to the Planetary Data System archive of NASA (https://pds.nasa.gov/).

We analysed the IOM samples used as references with the COSIMA ground model. The samples were prepared by demineralization of a sample of the Orgueil and the Murchison meteorites with hydrofluoric acid and hydrochloric acid[34]. This explains the observation of fluoride-bearing fragments, such as CF^+ at $m/z = 31.00$. The demineralization was not complete; we observe some other fragments, such as Si^+, Ca^+ and Fe^+ at $m/z = 27.98$, 39.96 and 55.93, respectively.

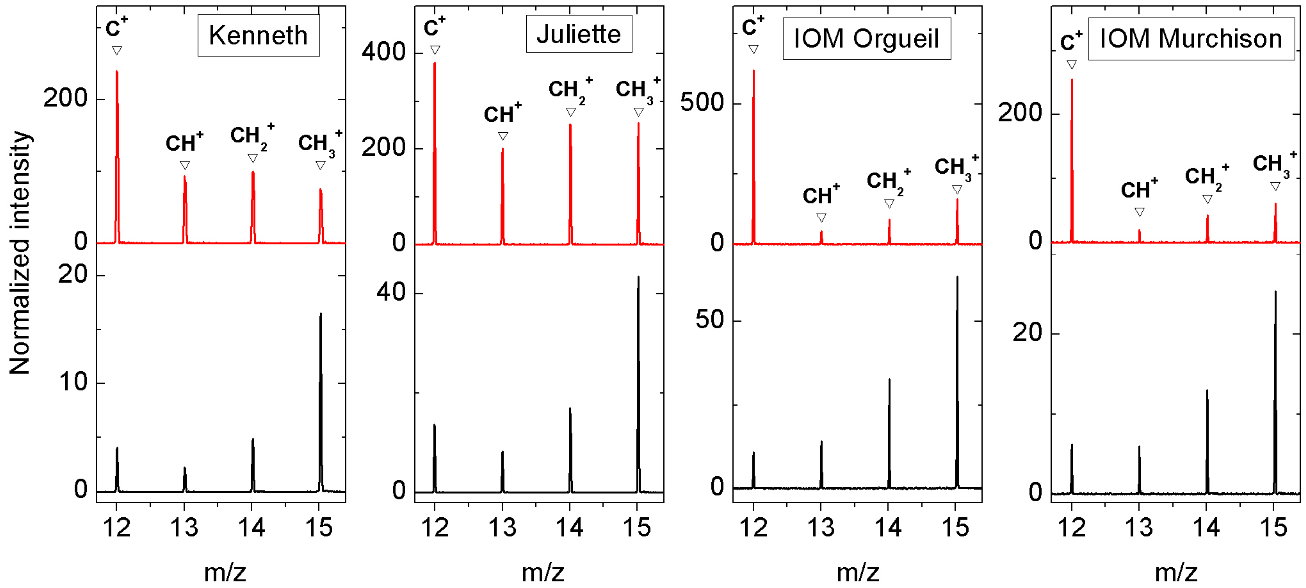
31. Brownlee, D. *et al.* Comet 81P/Wild 2 under a microscope. *Science* **314,** 1711–1716 (2006).
32. Kissel, J., Krueger, F. R., Silén, J. & Clark, B. C. The cometary and interstellar dust analyzer at comet 81P/Wild 2. *Science* **304,** 1774–1776 (2004).
33. Hornung, K. *et al.* Collecting cometary dust particles on metal blacks with the COSIMA instrument onboard ROSETTA. *Planet. Space Sci.* **103,** 309–317 (2014).
34. Kerridge, J. F., Chang, S. & Shipp, R. Isotopic characterisation of kerogen-like material in the Murchison carbonaceous chondrite. *Geochim. Cosmochim. Acta* **51,** 2527–2540 (1987).



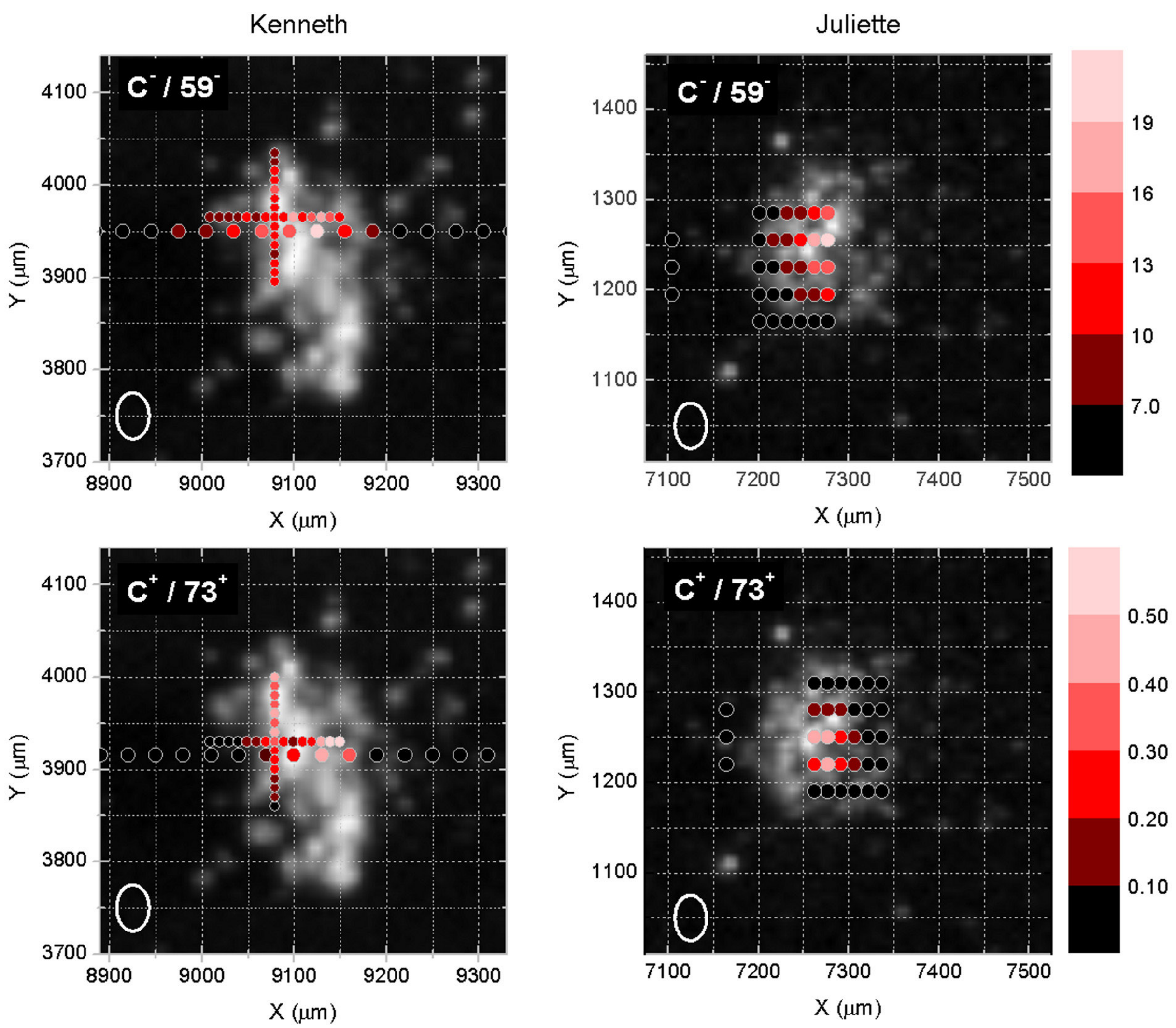
Extended Data Figure 1 | Optical images of the particles Kenneth and Juliette. These sub-pixel sampled images have an equivalent resolution of 10 μm (ref. 15). Each is the sum of two images, obtained with two grazing incidence illuminations from the left and the right. A square-root scaling has been used to bring out weakly illuminated regions. These images were acquired on 4 June 2015 (Kenneth) and 25 November 2015 (Juliette).



Extended Data Figure 2 | TOF-SIMS spectra acquired on five different cometary particles. The spectra in red were acquired on the cometary particles named Fadil, Jean-Pierre, Jessica, Kerttu and Stefane. The spectra in black were acquired on the respective gold targets, near the cometary particles. The latter spectra are normalized to the intensity of characteristic fragments of PDMS observed on the particles (see Methods). **a**, Positive-ion spectra. **b**, **c**, Enlargements from **a**. **d**, Negative-ion spectra. More details about the particles are in Extended Data Table 1.



Extended Data Figure 3 | Signatures of carbon in positive-ion spectra. The spectra in red were measured on the two cometary particles studied here, Kenneth and Juliette, and on an IOM sample extracted from the Orgueil and Murchison chondrites. The spectra in black were measured on the targets, near the cometary particles or the IOM samples.



Extended Data Figure 4 | Spatial distribution of the carbon secondary ions of the particles Kenneth and Juliette. The colour scale shows the values of the carbon/PDMS intensity ratios, on and off the particles, in negative and positive modes. The white ellipses indicate the size of the footprint of the primary ion beam, $35 \times 50\,\mu m^2$ (full-width at half-maximum). These spatial distributions have been superimposed on the optical images of the particles.

Extended Data Table 1 | Characteristics of the seven particles presented here

Particle Name	Starting date of the collection period	Total exposure time (days)	Analysis date in positive mode	Analysis date in negative mode	Typology of the particles
Kerttu	18/10/2014	6.62	26/03/2016 & 05/05/2016	26/03/2016 & 04/05/2016	C
Jean-Pierre	24/01/2015	1.02	08/01/2016 & 04/02/2016	08/01/2016 & 04/02/2016	S
Jessica	26/01/2015	0.97	07/01/2016 & 04/02/2016	06/01/2016 & 04/02/2016	S
Kenneth	11/05/2015	0.76	18/06/2015 & 02/07/2015	17/06/2015 & 01/07/2015	R
Juliette	23/10/2015	6.46	18/11/2015	18/11/2015	G
Fadil	16/11/2015	1.61	26/11/2015	25/11/2015	G
Stefane	17/01/2016	0.86	11/02/2016	11/02/2016	R

The table shows when the particles were collected (the starting date of the collection period and the total exposure time), analysis dates in positive and negative modes, and typologies (referring to the classification of ref. 15): C, compact particles; S, shattered clusters; G, glued clusters; R, rubble piles.

LETTER

doi:10.1038/nature19099

Nodal-chain metals

Tomáš Bzdušek[1], QuanSheng Wu[1,2], Andreas Rüegg[1], Manfred Sigrist[1] & Alexey A. Soluyanov[1,2,3]

The band theory of solids is arguably the most successful theory of condensed-matter physics, providing a description of the electronic energy levels in various materials. Electronic wavefunctions obtained from the band theory enable a topological characterization of metals for which the electronic spectrum may host robust, topologically protected, fermionic quasiparticles. Many of these quasiparticles are analogues of the elementary particles of the Standard Model[1–10], but others do not have a counterpart in relativistic high-energy theories[11–18]. A complete list of possible quasiparticles in solids is lacking, even in the non-interacting case. Here we describe the possible existence of a hitherto unrecognized type of fermionic excitation in metals. This excitation forms a nodal chain—a chain of connected loops in momentum space—along which conduction and valence bands touch. We prove that the nodal chain is topologically distinct from previously reported excitations. We discuss the symmetry requirements for the appearance of this excitation and predict that it is realized in an existing material, iridium tetrafluoride (IrF_4), as well as in other compounds of this class of materials. Using IrF_4 as an example, we provide a discussion of the topological surface states associated with the nodal chain. We argue that the presence of the nodal-chain fermions will result in anomalous magnetotransport properties, distinct from those of materials exhibiting previously known excitations.

Recently discovered Dirac and Weyl semimetals[3–10] host topologically protected degeneracy of four and two electronic bands, respectively, at isolated points in the Brillouin zone close to the Fermi level (E_F)[2,4,19]. The low-energy excitations in these materials are described by Dirac or Weyl Hamiltonians, as appropriate, of the relativistic quantum field theory, leading to the realization of the chiral anomaly[20,21] and topological surface Fermi arcs[3,9,10].

Owing to weaker symmetry constraints, condensed matter systems can realize quasiparticles that have no analogues in high-energy theories[11–18], hosting new physical phenomena. For example, in the presence of spin–orbit coupling, a valence and a conduction band with different mirror eigenvalues can touch along lines in mirror-invariant planes of the Brillouin zone, forming a so-called accidental nodal loop (ANL). The ANL materials are predicted to host special 'drumhead' surface states[18], which were argued to provide a route to higher-temperature superconductivity[22,23].

The spectrum of a nodal-chain fermion described here is illustrated in Fig. 1. The nodal chain consists of nodal loops, which are distinct from ANLs in that they are guaranteed to appear in the vicinity of the Fermi level (E_F) in certain non-centrosymmetric materials provided that their crystal structure has a non-symmorphic glide-plane symmetry $g=\{\sigma|\boldsymbol{t}\}$, formed by a reflection σ, followed by a translation by a fraction of a primitive lattice vector, $\boldsymbol{t}$. For several space groups listed in Fig. 1, such non-symmorphic nodal loops (NSNLs) appear on mutually orthogonal high-symmetry planes, touching each other at isolated points on a high-symmetry axis. Thus, a chain of double degeneracy is formed that goes across the entire Brillouin zone.

We first describe the building blocks of nodal chains—NSNLs. For spin–orbit coupled systems, $g^2=-\,e^{-i2\boldsymbol{k}\cdot\boldsymbol{t}_\parallel}$, where $\boldsymbol{k}$ is the electron momentum and $\boldsymbol{t}_{||}$ is the in-plane component of $\boldsymbol{t}$; consequently, the possible eigenvalues of g are $\eta_\pm(\boldsymbol{k})=\pm ie^{-i\boldsymbol{k}\cdot\boldsymbol{t}_\parallel}$, which are $\boldsymbol{k}$-dependent whenever $\boldsymbol{t}_{||}\neq \boldsymbol{0}$ (ref. 24).

The relation $\boldsymbol{\Gamma}_i\cdot\boldsymbol{t}_{||}=0$ (mod $\pi/2$) holds for any of the four in-plane time-reversal invariant momenta (TRIMs) $\boldsymbol{\Gamma}_i$, defined as $\boldsymbol{\Gamma}_i=-\boldsymbol{\Gamma}_i+\boldsymbol{G}$, with $\boldsymbol{G}$ a reciprocal lattice vector (see Supplementary Information). This definition makes it possible for the two TRIMs $\boldsymbol{\Gamma}_{1,2}$ to satisfy

$$(\boldsymbol{\Gamma}_1-\boldsymbol{\Gamma}_2)\cdot\boldsymbol{t}_{||}=\frac{\pi}{2}\ \mathrm{mod}\ \pi \tag{1}$$

so that the glide eigenvalues $\eta_\pm(\boldsymbol{k})$ are $\pm i$ at $\boldsymbol{k}=\boldsymbol{\Gamma}_1$ and ± 1 at $\boldsymbol{k}=\boldsymbol{\Gamma}_2$. Hence, along any in-plane path p that connects $\boldsymbol{\Gamma}_1$ to $\boldsymbol{\Gamma}_2$, the glide eigenvalues $\eta_\pm(\boldsymbol{k})$ must smoothly evolve from $(+i,-i)$ to $(+1,-1)$, as illustrated in Fig. 2a. However, in time-reversal-symmetric (Θ-symmetric) systems (see Supplementary Information for a generalization to antiferromagnetic systems), the bands form Kramers pairs, which are degenerate at TRIMs and carry complex-conjugate eigenvalues. Because the eigenvalues are no longer complex conjugates at $\boldsymbol{\Gamma}_2$, they belong to different Kramers doublets, meaning that there are several Kramers pairs that switch partners along p, as shown in Fig. 2b. This argument holds for any in-plane path p, and so there exists a nodal

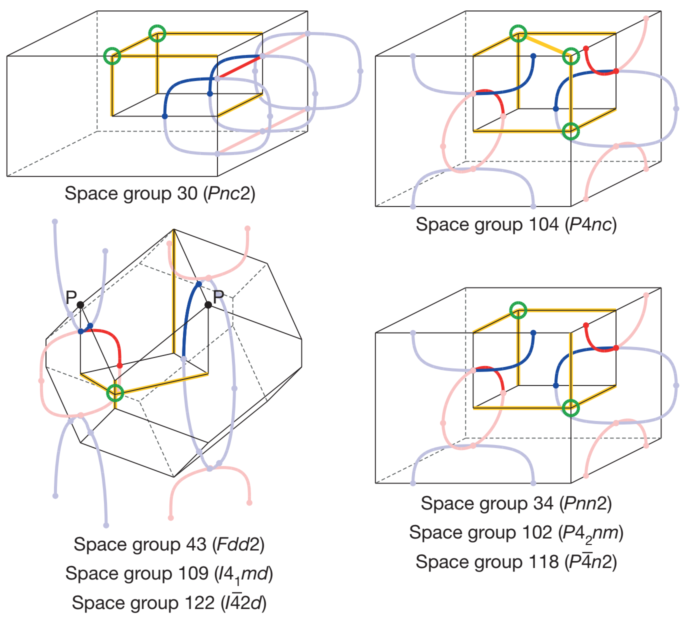


Figure 1 | Catalogue of nodal-chain metals. A nodal chain appears in metals with the space groups shown whenever there are $4n+2$ electrons per primitive unit cell. The blue and red lines show the nodal loops located in mutually orthogonal planes in the Brillouin zone. The additional double Weyl points are marked with green circles. The high-symmetry lines supporting a twofold degeneracy of valence (conduction) bands are highlighted in orange. In space groups 109 and 122 (shown bottom left), the touching point of the nodal loops is at the point P. The space groups are grouped according to their spectrum degeneracies at the high-symmetry points and lines.

[1]Institut für Theoretische Physik, ETH Zurich, 8093 Zurich, Switzerland. [2]Station Q Zurich, ETH Zurich, 8093 Zurich, Switzerland. [3]Department of Physics, St Petersburg State University, St Petersburg 199034, Russia.

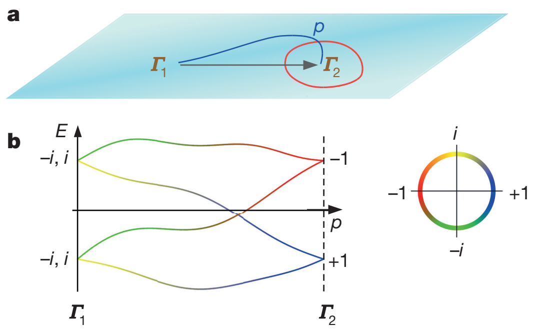


Figure 2 | Non-symmorphic nodal loop. a, Any path p connecting a pair of time-reversal-invariant momenta $\boldsymbol{\Gamma}_{1,2}$ in a glide-invariant plane (blue) that fulfil the criterion of equation (1) must have a gap closing point, which belongs to a non-symmorphic nodal loop (red). **b**, Two Kramers pairs shown along any path p connecting $\boldsymbol{\Gamma}_{1,2}$. The evolution of the glide eigenvalues along the path is shown in colour for all bands.

loop (the NSNL) separating the two TRIMs, shown as a red loop in Fig. 2a. A similar glide-plane argument plays a crucial role in realizing the hour-glass fermions[12] on the surfaces of certain insulators, but here we describe a three-dimensional metallic bulk excitation.

The illustration in Fig. 2b shows that NSNLs appear in materials in which bands come in quadruplets. Therefore, the NSNL is formed between valence and conduction bands whenever the number of electrons per unit cell is

$$\nu_{\text{filled}} = 4n + 2, \quad n \in \mathbb{N} \tag{2}$$

irrespective of all further material details. (Material examples of NSNLs formed by valence or conduction bands, and the ways in which NSNLs can be tuned to E_F, are discussed in Supplementary Information.)

The topological characterization and the existence of the drumhead surface states[18] is similar for ANLs and NSNLs (see Supplementary Information). Despite this similarity, we argue that low-energy excitations produced by ANLs and NSNLs are intrinsically distinct. Unlike ANLs, NSNLs are enforced by the symmetry of the underlying crystal structure. Moreover, NSNLs in Θ-symmetric, non-centrosymmetric systems always enclose a TRIM, and so a single nodal loop contains a time-reversed image of each Bloch state in addition to the state itself. In fact, if inversion-symmetry-breaking terms are smoothly tuned to zero in a NSNL Hamiltonian, then the NSNL shrinks into a Dirac point[4,19] (see Supplementary Information). This feature has immediate consequences in electron transport.

In particular, as outlined in Supplementary Information, application of a magnetic field in the direction orthogonal to the NSNL results in field-driven topological phase transitions. We find that the Landau levels of the conduction and valence bands touch at certain values B_c of the magnetic field, resulting in pumping of charge (equivalent to $e/2$ per area covered by a magnetic flux quantum, where e is the elementary charge) to the surface of the sample that is parallel to the plane of the NSNL. Hence, a step change in the Hall response of the metallic surface state is expected for magnetic field values B_c.

The response of the NSNLs to the mirror-symmetry-breaking, in-plane magnetic field is distinct from that of the ANLs. Although the Landau spectrum is gapped[25] for ANLs, it is always gapless for NSNLs. The crossing of the two Landau levels is protected by the product symmetry $\Theta \circ g$ that survives the application of the in-plane field. The gapless structure of the Landau levels suggests unusual transport properties for materials hosting NSNLs when an electric field is aligned with the in-plane magnetic field, similar to case of the chiral anomaly in Weyl and Dirac semimetals[20,21]. This dependence of the response on the direction of the magnetic field distinguishes NSNLs from all other known topological excitations.

Having established the NSNLs, we can now address systems with two glide planes. Such systems can accommodate nodal chains formed by a pair of touching NSNLs located in mutually orthogonal planes, while the bands at the touching point are still only doubly degenerate.

The criteria for the occurrence of a nodal chain are: (1) the system has to be symmetric under two inequivalent glide planes $g_{1,2} = \{\sigma_{1,2}|\boldsymbol{t}_{1,2}\}$ such that the criterion of equation (1) is fulfilled for the two TRIMs $\boldsymbol{\Gamma}_{1,2}$, which are located on the intersection of the two glide-invariant planes, for both translation vectors $\boldsymbol{t}_{1,2}$; and (2) the two bands forming the chain must belong to two-dimensional representations at $\boldsymbol{\Gamma}_{1,2}$, which split into one-dimensional representations on the high-symmetry line connecting $\boldsymbol{\Gamma}_1$ and $\boldsymbol{\Gamma}_2$.

Out of the 230 space groups[26], those satisfying the above criteria for two mutually orthogonal glide planes are listed in Fig. 1. The space group number 110 ($I4_1cd$) is discussed separately in Supplementary Information. In all the cases shown in Fig. 1, we find that at least one additional point of fourfold degeneracy, formed by two Weyl points of opposite chirality, is present at a high-symmetry point on the boundary of the Brillouin zone.

A nodal chain represents a new topological excitation, distinct from a collection of NSNLs. To see this, first note that the two NSNLs that form a nodal chain cannot be separated. The argument provided above for the appearance of the NSNL guarantees that there must be an odd number of band crossings along the high-symmetry line connecting $\boldsymbol{\Gamma}_1$ and $\boldsymbol{\Gamma}_2$.

The non-trivial transport properties of the nodal chain can be inferred from the above analysis of NSNLs in magnetic fields (a detailed study of the transport properties will be reported elsewhere (T.B., Q.S.W., A.A.S., manuscript in preparation)), suggesting several distinct scenarios for the Landau-level spectrum. Here we proceed with the analysis of the topological surface states of nodal chains that we illustrate using a particular real material example.

We found the nodal-chain state in iridium tetrafluoride (IrF_4). The orthorhombic crystal structure of this compound belongs to space group number 43 ($Fdd2$). The primitive unit cell contains two formula units[27] so that the number of electrons satisfies equation (2). Each iridium site is surrounded by an octahedron of six fluorine atoms, four of which are shared with the neighbouring octahedra. The octahedra form a bipartite lattice as shown in Fig. 3a, b (see Supplementary Information for a detailed description of the crystal structure). The space group contains two mutually orthogonal glide planes: g_1 and g_2, formed by a reflection about the (100) and (010) plane, respectively, followed by a translation of (1/4, 1/4, 1/4) in the reduced coordinates.

Possible antiferromagnetic ordering with a Néel temperature of less than about 100 K was reported for IrF_4 in magnetic susceptibility measurements[27]. A paramagnetic phase is expected to occur at temperatures above the Néel temperature, which are still accessible for angle-resolved photoemission spectroscopy (ARPES). We first discuss the paramagnetic phase, in which the crystal symmetries and band filling guarantee the presence of a nodal chain corresponding to the bottom left scenario in Fig. 1.

To study paramagnetic IrF_4 we performed first-principles calculations as detailed in Supplementary Information. The obtained band structure is shown in Fig. 3c. We indeed find a nodal chain, plotted in Fig. 4a, consisting of two NSNLs in the (100) and (010) planes. Both NSNLs cross the chemical potential four times, resulting in topologically protected touching points between electron and hole pockets (arrows in Fig. 3d). Similar touchings of carrier pockets, although of different topological origin, were predicted for type-II Weyl semimetals[11] and ANLs[16,18,23]. These Fermi-surface touching points can be probed using soft X-ray ARPES, and have been argued to be important for potential higher-temperature superconducting phases[22,23,28].

The nodal chain produces non-trivial topological surface states on the (100) surface of IrF_4, as shown in Fig. 3e, f. The projection of the (010) NSNL ((100) NSNL) onto the surface Brillouin zone is a line (oval), shown dashed in Fig. 3f. Fermi arcs arise from the touching points of the Fermi pockets. For the projection of the (100) NSNL (region 1 in Fig. 3h), a single such arc produced by the drumhead state

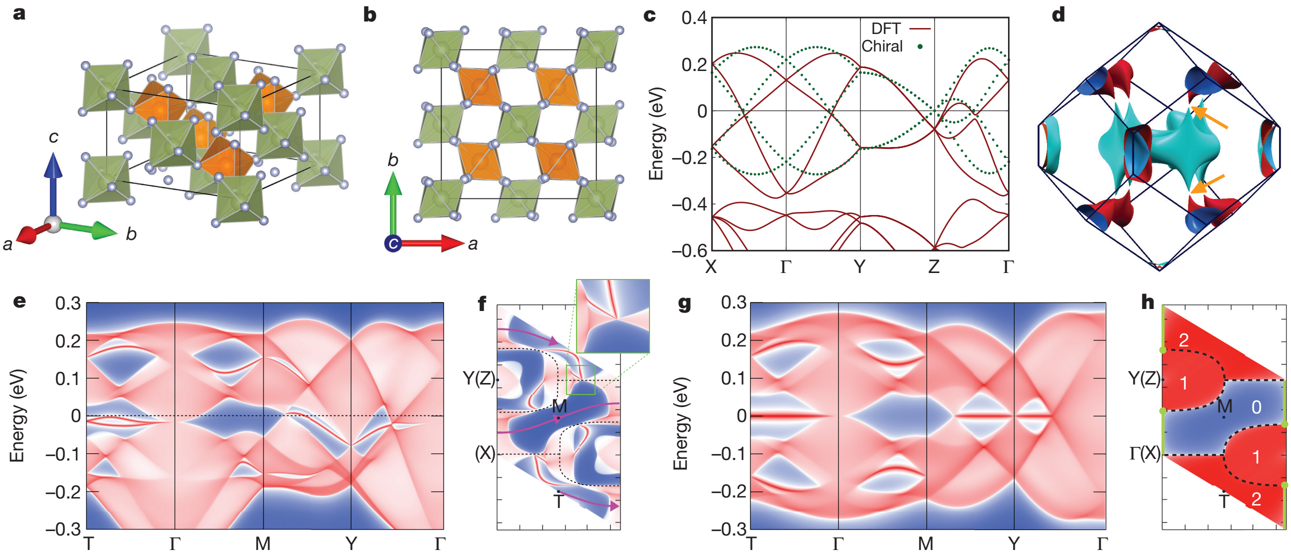


Figure 3 | Iridium tetrafluoride (IrF_4) and its band structure. a, The crystal structure of IrF_4. Corner-sharing octahedra of fluorine atoms enclose iridium atoms. The colour indicates the two sublattices related by an approximate chiral symmetry. **b**, The same structure viewed along the [001] axis. **c**, Band structure of paramagnetic IrF_4. Bands determined from first-principles (density functional theory, DFT; solid red lines) and from a tight-binding model with chiral symmetry (dotted green lines) are shown. **d**, The Fermi surface of IrF_4 consists of electron pockets (cyan) and hole pockets (red/dark blue when viewed from the outside/inside of the sheet), which touch (orange arrows) along the nodal rings. **e**, The density of states of the (100) surface shown along the high-symmetry lines of the surface Brillouin zone. Topological surface states are clearly visible. **f**, The density of states at the Fermi energy (dashed horizontal line in **e**), plotted in the (100) surface Brillouin zone. The end points of the surface Fermi arcs coincide with the projections of the bulk touching points of the electron and hole pockets. The dashed black line is the projection of the nodal chain into the surface Brillouin zone. The magenta line is the projection of a plane used to calculate the bulk $\mathbb{Z}_2$ invariant. **g**, **h**, Analogues of **e** and **f**, but for a tight-binding model with chiral symmetry. The numbers in **h** indicate the number of topological surface bands in that region; the green lines correspond to the projection of the additional nodal loop imposed by chiral symmetry into the surface Brillouin zone. The colours in **e**–**h** indicate the density of states, with blue corresponding to zero, white to intermediate and red to high density of states.

emerges from the touching point. However, the touching points that appear on a linear projection of the (010) NSNL produce two Fermi arcs, consistent with the fact that there are two such Fermi pocket touchings that project onto the same point in the surface Brillouin zone.

The arcs originating on different NSNLs are connected either directly or through a carrier pocket. Moreover, the $\mathbb{Z}_2$ invariant computed along the gapped, Θ-symmetric plane projected onto the magenta path in Fig. 3f is non-trivial. Hence, the path corresponds to an edge of a two-dimensional topological insulator, and has to host an odd number of Kramers pairs of edge states[29]. In accord with the observed connectivity of Fermi arcs, there is a single Kramers pair of such edge states (see Supplementary Information).

To understand why both Fermi arcs of the (010) NSNL appear on the same side of its projection onto the (100) surface, we need to expose the approximate chiral symmetry that is present in the material. We constructed a tight-binding model for the pseudospin-1/2 orbitals located on the iridium sites that represent the two sublattices of the IrF_4 structure, and fitted the parameters to reproduce the first-principles results (see Supplementary Information). We found that the avoided crossing along the Z–Γ line in Fig. 3c originates from the hoppings within the sublattices. The amplitudes of these hoppings are more than three times smaller than those of the inter-sublattice hoppings, meaning that there exists a weakly broken chiral symmetry in IrF_4, relating the two sublattices of the crystal structure. The chiral symmetry can be restored in the model by setting the intra-sublattice hoppings to zero. The corresponding band structure is shown in Fig. 3c, and it can be seen that the gap along the Z–Γ line now vanishes, and an additional nodal loop appears. It connects to the nodal chain, thus creating a nodal

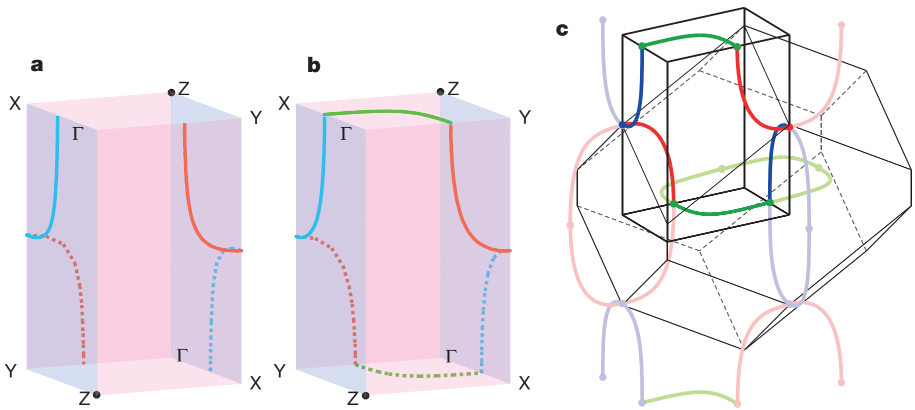


Figure 4 | Nodal chain and nodal net of IrF_4. a, Nodal-chain structure of IrF_4. **b**, Nodal net in the chiral-symmetric model of IrF_4. **c**, The form of the nodal net in extended ***k*** space. Different colours correspond to different orientations of the nodal loops. In **a** and **b**, the solid lines indicate nodal lines located in the visible (front and top) faces of the box and the dashed lines indicate nodal lines located in the hidden (bottom and back) faces of the box. In **c**, all lines are solid to highlight the net structure.

net, shown in Fig. 4. The projection of the additional nodal loop onto the (100) surface is shown in green in Fig. 3h.

Endowed with the chiral symmetry, the Hamiltonian allows for an additional topological classification (see Supplementary Information), which predicts two/one/zero surface modes to exist in the regions labelled 2/1/0 in Fig. 3h. In the presence of the chiral symmetry, all these regions are topologically distinct and separated by nodal loops. When the chiral symmetry is weakly broken in real IrF_4, only the parity of the number of surface states remains topologically protected and the additional nodal loop becomes gapped. However, because the breaking of the chiral symmetry is weak, the location of surface modes in the surface Brillouin zone of IrF_4 is inherited from the chiral-symmetric structure.

The possible antiferromagnetic ordering in IrF_4 at low temperatures preserves the nodal-chain structure if the magnetic moment is aligned with the [001] axis. In fact, the nodal chain survives weak breaking of time-reversal symmetry, but not the breaking of glide planes.

We also looked for other possible nodal chain candidates. Several reports[27,30] of stable XY_4 crystals (X = Ir, Ta, Re; Y = F, Cl, Br, I) with lattices formed of octahedra, similar to the IrF_4 lattice, exist, but with only fragmentary crystallographic data. Assuming these compounds crystallize in the same space group as IrF_4, we carried out an exhaustive first-principles study and found nodal chains in each of them (see Supplementary Information). We find that the particular shape of the chain and its position relative to the Fermi level depend on the lattice constants of the unit cell, suggesting the possibility of fine tuning with uniaxial or hydrostatic strains.

The prediction of the new nodal-chain state of matter in the IrF_4 class of materials opens up avenues for further study of novel physical properties associated with these compounds. The presence of both strongly and weakly correlated compounds in this family enables the interplay between the nodal-chain topology and electron–electron interactions, as well as magnetism, to be studied. The application of strains that break one of the glide planes in these compounds provides a route for a similar study of the NSNL phase, as well as for experimental probing of the anomalous magnetoelectric response predicted here for NSNLs.

Received 30 March; accepted 30 June 2016.
Published online 24 August 2016.

1. Volovik, G. E. *The Universe in a Helium Droplet* (Oxford Univ. Press, 2003).
2. Murakami, S. Phase transition between the quantum spin Hall and insulator phases in 3D: emergence of a topological gapless phase. *New J. Phys.* **9,** 356 (2007).
3. Wan, X., Turner, A. M., Vishwanath, A. & Savrasov, S. Y. Topological semimetal and Fermi-arc surface states in the electronic structure of pyrochlore iridates. *Phys. Rev. B* **83,** 205101 (2011).
4. Young, S. M. *et al.* Dirac semimetal in three dimensions. *Phys. Rev. Lett.* **108,** 140405 (2012).
5. Wang, Z. *et al.* Dirac semimetal and topological phase transitions in A_3Bi (A=Na, K, Rb). *Phys. Rev. B* **85,** 195320 (2012).
6. Liu, Z. K. *et al.* Discovery of a three-dimensional topological Dirac semimetal Na_3Bi. *Science* **343,** 864–867 (2014).
7. Weng, H., Chen, F. F., Fang, Z., Andrei Bernevig, B. & Dai, X. Weyl semimetal phase in noncentrosymmetric transition-metal monophosphides. *Phys. Rev. X* **5,** 011029 (2015).
8. Huang, S.-M. *et al.* A Weyl fermion semimetal with surface Fermi arcs in the transition metal monopnictide TaAs class. *Nat. Commun.* **6,** 7373 (2015).
9. Xu, S.-Y. *et al.* Discovery of a Weyl fermion semimetal and topological Fermi arcs. *Science* **349,** 613–617 (2015).
10. Lv, B. Q. *et al.* Experimental discovery of Weyl semimetal TaAs. *Phys. Rev. X* **5,** 031013 (2015).
11. Soluyanov, A. A. *et al.* Type-II Weyl semimetals. *Nature* **527,** 495–498 (2015).
12. Wang, Z., Alexandradinata, A., Cava, R. J. & Bernevig, B. A. Hourglass fermions. *Nature* **532,** 189–194 (2016).
13. Chang, P.-Y., Erten, O. & Coleman, P. Mobius Kondo insulators. Preprint at http://arxiv.org/abs/1603.03435 (2016).
14. Wieder, B. J., Kim, Y., Rappe, A. M. & Kane, C. L. Double Dirac semimetals in three dimensions. *Phys. Rev. Lett.* **116,** 186402 (2016).
15. Bradlyn, B. *et al.* Beyond Dirac and Weyl fermions: unconventional quasiparticles in conventional crystals. *Science* aaf5037 (2016).
16. Burkov, A. A., Hook, M. D. & Balents, L. Topological nodal semimetals. *Phys. Rev. B* **84,** 235126 (2011).
17. Fang, C., Chen, Y., Kee, H.-Y. & Fu, L. Topological nodal line semimetals with and without spin-orbital coupling. *Phys. Rev. B* **92,** 081201(R) (2015).
18. Weng, H. *et al.* Topological node-line semimetal in three-dimensional graphene networks. *Phys. Rev. B* **92,** 045108 (2015).
19. Yang, B.-J. & Nagaosa, N. Classification of stable three-dimensional Dirac semimetals with nontrivial topology. *Nat. Commun.* **5,** 4898 (2014).
20. Nielsen, H. B. & Ninomiya, M. The Adler–Bell–Jackiw anomaly and Weyl fermions in a crystal. *Phys. Lett. B* **130,** 389–396 (1983).
21. Huang, X. *et al.* Observation of the chiral-anomaly-induced negative magnetoresistance in 3D Weyl semimetal TaAs. *Phys. Rev. X* **5,** 031023 (2015).
22. Heikkilä, T. T., Kopnin, N. B. & Volovik, G. E. Flat bands in topological media. *JETP Lett.* **94,** 233–239 (2011).
23. Heikkilä, T. T. & Volovik, G. E. in *Basic Physics of Functionalized Graphite* (ed. Esquinazi, P. D.) 123–143 (Springer, 2016).
24. Young, S. M. & Kane, C. L. Dirac semimetal in two dimensions. *Phys. Rev. Lett.* **115,** 126803 (2015).
25. Rhim, J.-W. & Kim, Y. B. Landau level quantization and almost flat modes in three-dimensional semimetals with nodal ring spectra. *Phys. Rev. B* **92,** 045126 (2015).
26. Bradley, C. J. & Cracknell, A. P. *The Mathematical Theory of Symmetry in Solids* (Clarendon Press, 1972).
27. Rao, P. R., Tressaud, A. & Bartlett, N. The tetrafluorides of iridium, rhodium and palladium. *J. Inorg. Nucl. Chem.* **28** (Suppl.), 23–28 (1976).
28. Yudin, D. *et al.* Fermi condensation near van Hove singularities within the Hubbard model on the triangular lattice. *Phys. Rev. Lett.* **112,** 070403 (2014).
29. Kane, C. L. & Mele, E. J. Z_2 topological order and the quantum spin Hall effect. *Phys. Rev. Lett.* **95,** 146802 (2005).
30. Habermehl, K. *Neue Untersuchungen an Halogeniden des Niobs und Tantals*. PhD thesis, Univ. Köln, http://kups.ub.uni-koeln.de/3103/ (2010).

Supplementary Information is available in the online version of the paper.

Acknowledgements The crystal structures in Fig. 3 were plotted using VESTA 3. We thank A. Bouhon, C. L. Kane, G. E. Volovik, B. A. Bernevig and R. J. Cava for discussions. T.B., A.R. and M.S. acknowledge financial support through an ETH research grant and the Swiss National Science Foundation. Q.S.W. and A.A.S. acknowledge the support of Microsoft Research and the Swiss National Science Foundation through the National Competence Centres in Research MARVEL and QSIT.

Author Contributions T.B. initiated the project, carried out the theoretical analysis and determined the suitable space groups. Q.S.W. discovered the IrF_4 material class and performed the first-principle studies. T.B. and A.A.S. wrote the manuscript. All authors contributed to the theoretical discussion and the final version of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to T.B. (bzdusek@itp.phys.ethz.ch).

Reviewer Information *Nature* thanks E. Bergholtz, R. Nandkishore and the other anonymous reviewer(s) for their contribution to the peer review of this work.

doi:10.1038/nature19089

Surface patterning of nanoparticles with polymer patches

Rachelle M. Choueiri[1], Elizabeth Galati[1], Héloïse Thérien-Aubin[1], Anna Klinkova[1], Egor M. Larin[1], Ana Querejeta-Fernández[1], Lili Han[2,3], Huolin L. Xin[2], Oleg Gang[2]†, Ekaterina B. Zhulina[4,5], Michael Rubinstein[6] & Eugenia Kumacheva[1,7,8]

Patterning of colloidal particles with chemically or topographically distinct surface domains (patches) has attracted intense research interest[1–3]. Surface-patterned particles act as colloidal analogues of atoms and molecules[4,5], serve as model systems in studies of phase transitions in liquid systems[6], behave as 'colloidal surfactants'[7] and function as templates for the synthesis of hybrid particles[8]. The generation of micrometre- and submicrometre-sized patchy colloids is now efficient[9–11], but surface patterning of inorganic colloidal nanoparticles with dimensions of the order of tens of nanometres is uncommon. Such nanoparticles exhibit size- and shape-dependent optical, electronic and magnetic properties, and their assemblies show new collective properties[12]. At present, nanoparticle patterning is limited to the generation of two-patch nanoparticles[13–15], and nanoparticles with surface ripples[16] or a 'raspberry' surface morphology[17]. Here we demonstrate nanoparticle surface patterning, which utilizes thermodynamically driven segregation of polymer ligands from a uniform polymer brush into surface-pinned micelles following a change in solvent quality. Patch formation is reversible but can be permanently preserved using a photocrosslinking step. The methodology offers the ability to control the dimensions of patches, their spatial distribution and the number of patches per nanoparticle, in agreement with a theoretical model. The versatility of the strategy is demonstrated by patterning nanoparticles with different dimensions, shapes and compositions, tethered with various types of polymers and subjected to different external stimuli. These patchy nanocolloids have potential applications in fundamental research, the self-assembly of nanomaterials, diagnostics, sensing and colloidal stabilization.

We hypothesized that the segregation of polymer ligands into surface-pinned micelles having a footprint area comparable with the surface area of the nanoparticle could be used as a thermodynamically mediated strategy for the patterning of the high-curvature surface of nanocolloids. The formation of pinned micelles on planar surfaces has been studied for polymer molecules strongly grafted to a macroscopic planar surface[18–21]. When a polymer-tethered substrate was transferred from a good to a poor solvent, a smooth layer segregated into micelles composed of a dense core and stretched surface-tethered 'legs' (Fig. 1a, top).

The proposed approach to patchy nanoparticles is illustrated in Fig. 1a (bottom). On the nanoparticle surface, following the reduction in solvent quality, a uniformly thick polymer brush layer breaks up into a discrete number of pinned micelles (patches). The process is driven by attractive polymer–polymer interactions and the competition between the polymer grafting constraints and the reduction in its interfacial free energy.

Here we validate this approach for nanoparticles with different dimensions, shapes and chemical compositions, which were capped with various types of polymer and copolymer ligands and subjected to different external stimuli. We show experimentally and theoretically that the size of patches is governed by the polymer dimensions and grafting density, whereas the number of patches per nanoparticle is determined by the ratio between the nanoparticle diameter and polymer size. The patches could be permanently vitrified by polymer photocrosslinking. The resulting patchy nanoparticles acted as *in situ* colloidal surfactants and their self-assembly exhibited new binding modalities.

We note that in addition to the generation of patchy nanoparticles, polymer segregation on the surface of nanoparticles has other far-reaching implications. Polymer-tethered nanoparticles have a broad range of applications in imaging and medical diagnostics[22], therapeutics[23], and chemical sensing[24]. The change in morphology of the polymer layer under varying ambient conditions is of fundamental importance and can be used for the efficient design of nanoparticles aimed at specific applications.

To explore the proposed approach, we synthesized gold spherical nanoparticles (nanospheres) with a mean diameter D in the range from 20 ± 1.0 nm to 80 ± 1.5 nm, which were stabilized with cetyltrimethylammonium bromide or cetylpyridinium chloride. These low-molecular-mass ligands were replaced with thiol-terminated polystyrene molecules with a molecular mass of 29,000 g mol^{-1} or 50,000 g mol^{-1} (Supplementary Figs 1–4 and Supplementary Tables 1–3). (Henceforth we refer to these polymers as polystyrene-30K and polystyrene-50K, respectively.) The polymer-capped nanospheres were dispersed in dimethylformamide (DMF), a good solvent for polystyrene molecules (the value of the second virial coefficient, A_2, is 3.5×10^{-4} mol cm^3 g^{-2}, equivalent to a Flory–Huggins interaction parameter of 0.46)[25].

Figure 1b shows a transmission electron microscopy (TEM) image of 20-nm-diameter nanospheres functionalized with polystyrene-50K. When cast on the grid from the solution in DMF, the nanospheres were engulfed by a uniformly thick polymer shell. Following the reduction in solvent quality for the polystyrene ligands—by adding water to the nanosphere solution in DMF—the polymer layer transformed into a surface patch (Fig. 1c). Since the surface mobility of thiol-terminated molecules is suppressed for multi-facet gold nanospheres and for high-molecular-mass ligands, and since their lateral motion is generally slow[26], we expected that in a poor solvent, stretched polystyrene-50K molecules would be grafted to the nanosphere surface, as is shown in Fig. 1a, bottom. Upon polymer surface segregation, the yield of patchy nanospheres was 65%; other species included small self-assembled nanosphere clusters (32%) and nanospheres with a smooth shell (3%). After removal of the clusters by centrifugation, the fraction of patchy nanospheres was about 98%. Patch formation was reversible: upon dilution of the solution with DMF to a water concentration of $C_w < 1$ vol% the core–shell nanosphere morphology was recovered.

[1]Department of Chemistry, University of Toronto, 80 Saint George Street, Toronto, Ontario M5S 3H6, Canada. [2]Center for Functional Nanomaterials, Brookhaven National Laboratory, Upton, New York 11973, USA. [3]Institute of New Energy Materials, School of Materials Science and Engineering, Tianjin University, Tianjin 300072, China. [4]Institute of Macromolecular Compounds of the Russian Academy of Sciences, Saint Petersburg, 199004, Russia. [5]Saint Petersburg National University of Informational Technologies, Mechanics and Optics, Saint Petersburg, 197101, Russia. [6]Department of Chemistry, University of North Carolina, Chapel Hill, North Carolina 27599-3290, USA. [7]Institute of Biomaterials and Biomedical Engineering, University of Toronto, 4 Taddle Creek Road, Toronto, Ontario M5S 3G9, Canada. [8]Department of Chemical Engineering and Applied Chemistry, University of Toronto, 200 College Street, Toronto, Ontario M5S 3E5, Canada. †Present addresses: Department of Chemical Engineering, Columbia University, New York 10027, USA; Department of Applied Physics and Applied Mathematics, Columbia University, New York 10027, USA.

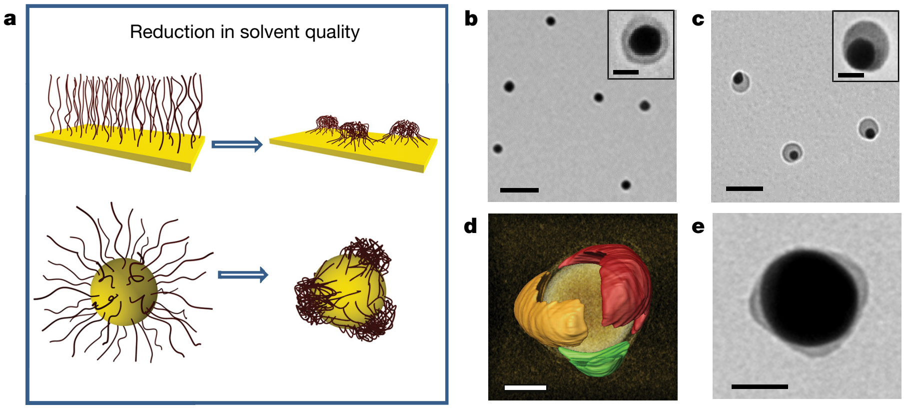


Figure 1 | Polymer segregation on the nanoparticle surface. **a**, Schematics of solvent-mediated formation of pinned polymer micelles (surface patches) on a planar macroscopic surface (top) and on the nanoparticle surface (bottom). **b**, **c**, TEM images of gold nanospheres capped with polystyrene-50K at the grafting density of 0.03 chains per square nanometre and deposited on the grid from the 0.3 nM nanosphere solution in DMF (**b**) and from the DMF/water mixture at $C_w = 4$ vol% after 24 h incubation at 40 °C (**c**). Scale bars in **b** and **c** are 100 nm. Insets in **b** and **c** show the corresponding images of individual nanospheres. Inset scale bars are 20 nm. **d**, Electron tomography reconstruction image of the 60-nm-diameter nanosphere with three polystyrene-50K patches, each shown for clarity with a different arbitrary colour. The image of a gold core is removed to highlight the structure of polymer patches. The estimated resolution is 2–3 nm. Patchy nanospheres were formed as described in **c**. The grafting density of polystyrene-50K is 0.02 chains per square nanometre. **e**, TEM image of the gold nanosphere carrying photocrosslinked thiol-terminated polystyrene-*co*-polyisoprene patches preserved after 24 h incubation in tetrahydrofuran (a good solvent for polystyrene-*co*-polyisoprene). Original patchy nanoparticles were formed and crosslinked in the DMF/water mixture at $C_w = 1$ vol%. Scale bars in **d** and **e** are 20 nm.

The formation of multi-patch nanospheres was explored for nanospheres with larger dimensions. Figure 1d shows a three-dimensional electron tomography reconstruction image of the 60-nm-diameter patchy gold nanosphere capped with polystyrene-50K (see Supplementary Information for details). The nanosphere carried three polymer patches, each shown with a different arbitrary colour for clarity. The side view, obtained from tomographic reconstruction, revealed an elongated patch shape, which could be induced by the partial wetting of the substrate with the polymer solution. Some accumulation of the polymer at the nanosphere–substrate interface (Supplementary Fig. 21, Supplementary Videos 1 and 2), supports this assumption.

To ensure that polymer surface segregation occurs in solution, patchy gold nanospheres tethered with thiol-terminated random copolymer polystyrene-*co*-polyisoprene were introduced into a 0.05 wt% solution of photoinitiator azobisisobutyronitrile in the DMF/water mixture and exposed to ultraviolet irradiation for 5 min. Partitioning of the photoinitiator into the patches and copolymer photocrosslinking yielded a permanent patchy structure on the nanosphere surface, which was preserved in tetrahydrofuran, a good solvent for the copolymer (Fig. 1e). Without crosslinking, the patches transformed into a smooth shell (Supplementary Fig. 20). Below, we refer to the non-crosslinked patchy nanoparticles, which were characterized by analysing their two-dimensional projections in TEM images.

Patch formation and their structure were governed by polymer length, nanosphere diameter, and polymer grafting density. In the first series of experiments, we examined transitions between the nanospheres with a smooth polymer shell (core–shell nanospheres) and patchy nanospheres at varying ratios between the nanosphere and polymer size. Nanospheres with diameter $20\,\text{nm} \leq D \leq 80\,\text{nm}$ were capped with polystyrene-30K or polystyrene-50K, having molecular radii, R, of 11 nm or 15 nm, respectively[27]. (The unperturbed polymer chain size is typically defined as the root-mean-square end-to-end distance R of a polymer in its ideal conformation, which we call here the 'chain radius' or 'molecular radius' R.) At a constant grafting density σ of polymer chains with a radius R, polymer segregation was favoured for small nanospheres (Fig. 2a, top, and Supplementary Figs 9–13 and 22–24), while the reduction in R led to a larger number of patches per nanosphere at constant D and σ (Fig. 2a, bottom). Figure 2b illustrates these trends for varying D/R ratios (characterizing a different extent of stretching of the micellar 'legs'). For example, at $D/R = 1.3$ ($D = 20$ nm, polystyrene-50K), 98% of the nanospheres had a single patch, while at $D/R = 2.6$ ($D = 34$ nm, polystyrene-50K), 34% and 53% of the nanospheres had one and two patches, respectively. The angles between the patch centres were $170° \pm 9°$ and $120° \pm 13°$ for two-patch and three-patch nanospheres, correspondingly, characterizing the uniformity of patch distribution on the nanosphere surface. The average maximum patch height increased with polymer molecular mass: for two-patch 40-nm-diameter nanospheres capped with polystyrene-30K and polystyrene-50K the patch height was 6.5 ± 0.65 nm and 9.0 ± 0.31 nm, respectively.

Next, the formation of patches was examined while varying the grafting density, σ, of polystyrene-50K capping nanospheres with different dimensions. Figure 2c shows the experimental diagram of nanosphere states, plotted in D–σ parameter space. The transition between the core–shell and patchy nanospheres was favoured at decreasing σ and D (or increasing curvature) values, signified by the negative slope of the boundary solid blue line. In the patchy region, the average number of patches per nanosphere, n, increased with the nanosphere diameter and did not noticeably vary with σ (and was thus averaged over the range of σ studied). The latter effect further supports the lack of lateral mobility of thiol-terminated polymer ligands on the nanosphere surface. The polymer grafting density influenced patch dimensions. For example, for three-patch nanospheres with $D = 60$ nm the average maximum height of the polystyrene-50K patch decreased from 7.7 ± 1.1 nm to 3.1 ± 0.35 nm when σ reduced from 0.02 to 0.003 chains per square nanometre, respectively.

The trends shown in Fig. 2c were captured in the theoretical state diagram in Fig. 2d (the theoretical model is described in Supplementary Information). The structure of the polymer layer on the nanosphere surface was governed by the polymer–solvent interfacial energy and the energy of stretching of end-tethered polymer molecules. In Fig. 2d, at high σ values (the right region of the diagram), extended polymer chains minimized their interfacial and stretching energies by forming a smooth layer[19]. At lower values of σ (the left side of the diagram), the layer became thinner than the unperturbed molecular size of the polymer and the interfacial polymer–solvent energy was lowered by polymer segregation in pinned micelles. The elastic energy of stretched micellar 'legs' was comparable to the polymer–solvent interfacial energy. For large nanospheres, the transition between the two regions is shown as a blue line approaching the

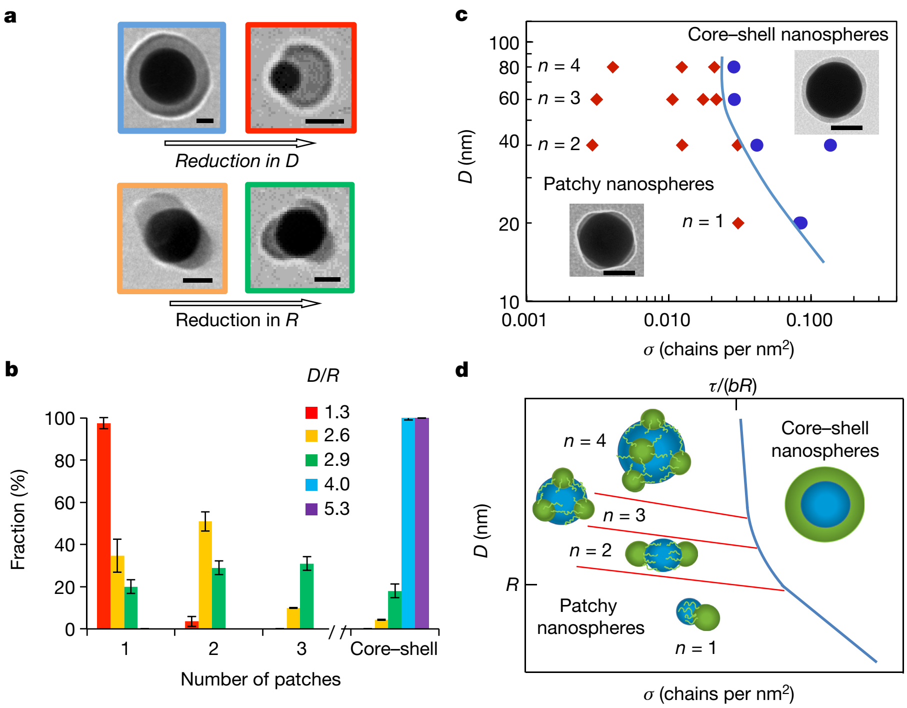


Figure 2 | Structural transitions in the polymer layer on the surface of gold nanospheres. a, Effect of nanosphere size (top) and polymer dimensions (bottom) on patch formation. The nanospheres are functionalized with polystyrene-50K (top row and bottom left) and polystyrene-30K (bottom right) at $\sigma = 0.03$ chains per square nanometre. Scale bars are 25 nm. **b**, Distribution of populations of nanospheres with a different patch number. The red, yellow, blue and violet bars correspond to the 20-, 40-, 60- and 80-nm-diameter nanospheres capped with polystyrene-50K, respectively; the green bar represents 32-nm-diameter nanospheres functionalized with polystyrene-30K. $\sigma = 0.03$ chains per square nanometre. The error bars represent the standard deviations. Each experiment was run in triplicate. The inset shows the D/R ratios, with colours corresponding to the colours of bars and the frames of the images in **a**. **c**, Experimental diagram of nanosphere states. The blue line separates the regions of core–shell and patchy nanospheres with a different average patch number n. The insets illustrate patchy and a core–shell nanospheres with σ of 0.012 and 0.03 chains per square nanometre, respectively. Scale bars are 50 nm. In **b** and **c**, 200–300 nanospheres were analysed for each population. **d**, Theoretical diagram of nanosphere states. Transitions between nanospheres with different values of n begin at $D \approx R$ (red lines). The blue line shows the boundary between the smooth and patchy polymer layer, approaching the grafting density $\sigma = \tau/(bR)$ for large nanospheres, similar to **c**.

grafting density $\tau/(bR)$, where τ accounts for the solvent quality and b is the monomer length.

The competition between the interfacial and stretching energies resulted in the optimized micelle footprint area $A \approx (N^2\tau/\sigma)^{2/5}$, where N is the polymer degree of polymerization. Since the number of micelles per nanosphere is proportional to the ratio between the nanosphere surface area πD^2, and the micelle footprint area A, for varying nanosphere dimensions and/or polymer grafting densities, transitions were expected between the nanospheres with n and $n+1$ patches. The inclined red lines with constant $\pi D^2/A$ ratios in Fig. 2d outline these transitions, with single-patch ($n = 1$) nanospheres at the bottom and a transition between $n = 1$ and $n > 1$ at $D \approx R$.

The effect of nanosphere size (or surface curvature) on patch formation was revealed by the position and incline of the boundary between the core–shell and patchy nanosphere states. The balance between the interfacial energy of the polymer and the free energy of stretching of the micellar 'legs' led to a higher stability of micelles on small nanospheres and hence a negative slope of the boundary line. Thus overall, the experimental and theoretical results were in excellent agreement.

The versatility of the nanopatterning method was explored for nanoparticles with different shapes and compositions, capped with different polymer ligands strongly binding to the nanoparticle surface and subjected to different solvents (Supplementary Figs 5–8 and 14–19). Following the prediction of the theoretical model on a stronger tendency for patch formation on surfaces with a high curvature, we examined polymer segregation on nanorods, nanocubes and triangular nanoprisms. Following incubation of polystyrene-50K-capped spherocylindrical gold nanorods in the DMF/water mixture, a uniform polymer layer separated into two distinct patches engulfing the nanorod tips (Fig. 3a). Similar polymer segregation towards metal tips occurred for gold nanorods with a dumbbell shape (Fig. 3b). Patches of polystyrene-50K formed on the edges of silver nanocubes and triangular nanoprisms incubated in a poor solvent (Fig. 3c, d), as well as on edges of gold nanocubes in the tetrahydrofuran/water mixture (Supplementary Fig. 18).

Other polymer ligands exhibited qualitatively similar surface segregation in a poor solvent. Thiol-terminated poly(4-vinyl pyridine) on gold nanospheres formed into a patch following an increase in pH of an aqueous nanosphere solution from 2.5 to 11.5, at which the polymer became hydrophobic (Fig. 3e). Thiol-terminated poly(N-vinyl carbazole) ligands split into two patches on the surface of gold nanospheres incubated in the DMF/water mixture (Fig. 3f).

The generation of patchy nanoparticles enabled preliminary exploration of their new self-assembly modalities. In the present work, to produce individual patchy nanoparticles, we suppressed their self-assembly in a poor solvent by using dilute solutions; however, given sufficient time, patchy nanospheres assembled in chains co-existing with small clusters of two to three single-patch nanospheres (Supplementary Fig. 25). We isolated nanosphere dimers (shown in Fig. 4a) by centrifugation of the colloidal solution and separation of larger nanosphere assemblies. The ability to control the separation between the gold surfaces by varying polymer grafting density or molecular mass enables control over hot spots of a strong electric field in the gap between the nanospheres in the dimers, making them useful in Raman scattering[28]. Inspection of isolated chains of patchy nanospheres revealed that they were built from dimers and trimers (Fig. 4b and Supplementary Figs 25 and 26), suggesting a sequential mechanism of the self-assembly of patchy

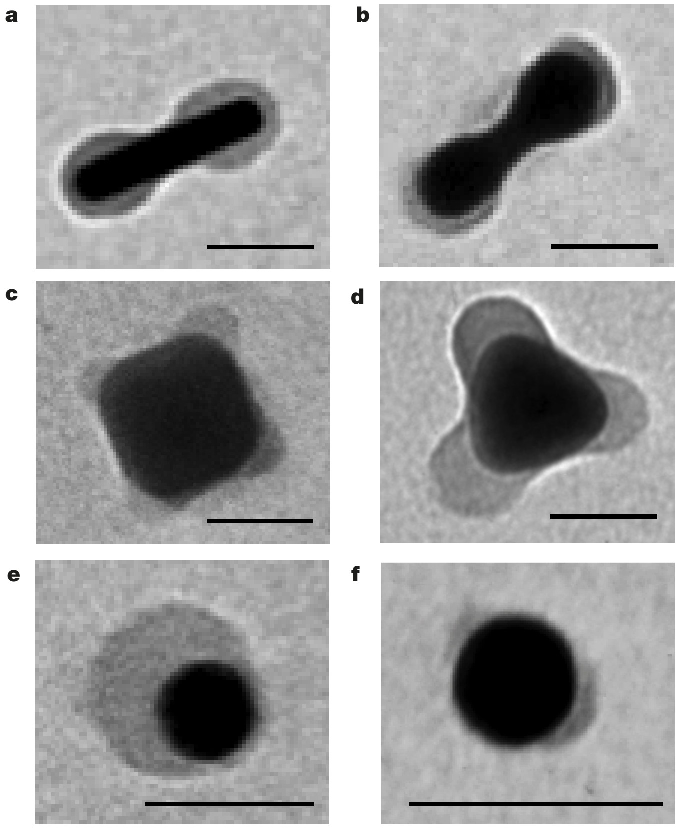


Figure 3 | Generality of polymer patterning of nanoparticle surface. **a**–**d**, TEM images of polystyrene-50K-coated gold spherocylindrical nanorods (**a**), gold dumbbell-shaped nanorods (**b**), silver nanocubes (**c**) and silver triangular prisms (**d**), all in the DMF/water mixture at $C_w = 4$ vol%. **e**, **f**, TEM images of thiol-terminated poly(4-vinyl pyridine) ($M_n = 22{,}000\,\mathrm{g\,mol^{-1}}$) in water at pH = 10.5 (**e**) and thiol-terminated poly(*N*-vinylcarbazole) ($M_n = 19{,}800\,\mathrm{g\,mol^{-1}}$) in the DMF/water mixture at $C_w = 4$ vol% (**f**), both on the surface of gold nanospheres. All scale bars are 40 nm.

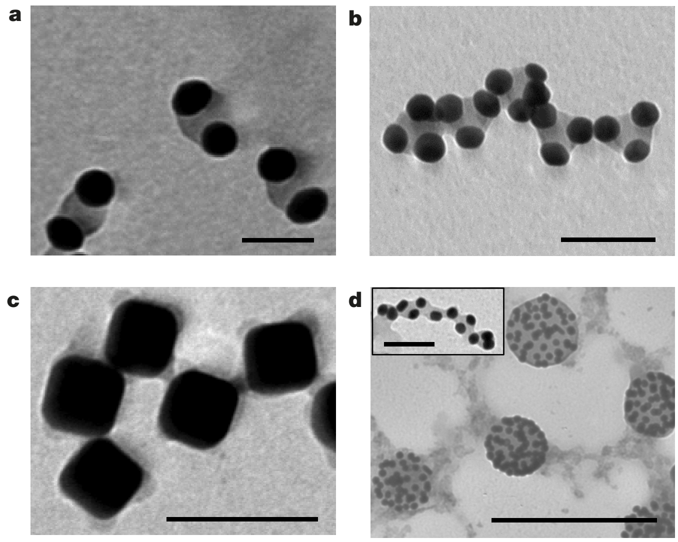


Figure 4 | Self-assembly of patterned nanoparticles. **a**, Dimers of single-patch gold nanospheres. **b**, Self-assembly of trimers of single-patch gold nanospheres in chains. In **a** and **b** the nanospheres were capped with polystyrene-50K and incubated for 15 days in the DMF/water solution at $C_w = 4$ vol% at 40 °C. Scale bars in **a** and **b** are 40 nm. **c**, Self-assembly of patchy silver nanocubes functionalized with polystyrene-50K in the DMF/water mixture at $C_w = 20$ vol%; scale bar is 100 nm. **d**, Self-assembly of gold nanospheres on the surface of droplets enriched with free non-thiolated polystyrene. The self-assembly was induced by adding water at $C_w = 4$ vol% to the mixed solution of free non-thiolated polystyrene ($M_n = 50{,}000\,\mathrm{g\,mol^{-1}}$) and gold nanospheres tethered with polystyrene-50K in DMF; scale bar is 250 nm. The inset to **d** shows self-assembly of patchy polystyrene-50K-capped gold nanospheres in the DMF/water mixture at $C_w = 4$ vol% in the presence of 0.625 nM of non-thiolated polystyrene, following 5 min sonication of the solution. Scale bar is 40 nm.

nanospheres: a faster assembly of dimers and trimers and a slower assembly of these building blocks in chains, in comparison with the self-assembly of non-patchy nanospheres[29].

A new binding modality was also observed for patchy nanocubes undergoing self-assembly in an open, 'checkerboard' structure, owing to the binding of nanocube edges in a poor solvent (Fig. 4c), markedly different from the face-to-face assembly of the nanocubes uniformly coated with polystyrene ligands (Supplementary Fig. 27). For patchy nanocubes, the face-to-face and the 'checkerboard' assembly via the formation of four bonds between the edges may result in a similar reduction in the surface free energy of the system, whereas for non-patchy nanocubes, the formation of close-packed structures would be favoured, owing to the maximum screening of unfavourable polymer interactions with a poor solvent[30].

The amphiphilic nature of patchy nanospheres led to their assembly at the interface between immiscible liquids, thus reducing the surface energy of the system and behaving as colloidal surfactants. Following the addition of water to the mixture of polystyrene-capped gold nanospheres and non-thiolated free polystyrene molecules in DMF, the reduction in solvent quality led to the formation of patchy nanospheres and polystyrene-rich droplets. The nanospheres self-assembled on the droplet surface, with a polystyrene patch immersed in the droplet (Fig. 4c). A considerably higher energy of attachment of patchy nanospheres to liquid–liquid interface[6], in comparison with conventional Pickering emulsions, is expected to provide enhanced stabilization properties of emulsions. Sonication of patchy nanospheres and non-thiolated polystyrene in the DMF/water solution led to the formation of elongated polystyrene species decorated with patchy nanospheres (inset to Fig. 4d).

We have thus developed a new strategy for nanoparticle surface patterning that is governed by thermodynamically controlled segregation of polymer ligands in pinned micelles with a footprint area comparable with the nanoparticle surface area. Polymer segregation is favoured for small nanoparticles with a large surface curvature. The experimental results were in excellent agreement with the proposed theoretical model. The described patterning strategy can be used for the generation of reconfigurable nanocolloids: reversible transitions between a smooth polymer shell and surface patches can be triggered by illumination, change in temperature, ionic strength or pH of the solution, that is, the stimuli changing the solvent quality. On demand, polymer patches can be 'locked' by permanent crosslinking, which would suppress nanoparticle assembly[31] and enable the utilization of solutions with a higher nanoparticle concentration, thereby increasing the yield of patchy nanoparticles.

The utilization of block copolymers will facilitate nanoparticle patterning with a variety of pinned micelle structures, including co-micelles, which may tailor new functionalities to patchy nanoparticles. 'Grafting-from' surface functionalization[32] and fractionation of nanoparticles with a particular number of patches will enhance control over the number of patches per nanoparticle. Patterning of multicomponent nanoparticles and the self-assembly of patterned nanoparticles into complex, hierarchical structures are other directions to explore. Furthermore, given the remarkable progress in the synthesis of nanoparticles with different shapes, the proposed strategy enables fundamental studies of polymer segregation on surfaces with large curvatures or surfaces with multiple curvatures.

Received 20 January; accepted 28 June 2016.
Published online 24 August; corrected online 5 October 2016 (see full-text HTML version for details).

1. Bianchi, E., Blaak, R. & Likos, C. N. Patchy colloids: state of the art and perspectives. *Phys. Chem. Chem. Phys.* **13,** 6397–6410 (2011).
2. Preisler, Z., Vissers, T., Munaò, G., Smallenburg, F. & Sciortino, F. Equilibrium phases of one-patch colloids with short-range attractions. *Soft Matter* **10,** 5121–5128 (2014).
3. Chen, Q., Bae, S. C. & Granick, S. Directed self-assembly of a colloidal kagome lattice. *Nature* **469,** 381–384 (2011).
4. Glotzer, S. C. & Solomon, M. J. Anisotropy of building blocks and their assembly into complex structures. *Nat. Mater.* **6,** 557–562 (2007).
5. Gröschel, A. H. *et al.* Guided hierarchical co-assembly of soft patchy nanoparticles. *Nature* **503,** 247–251 (2013).
6. Kern, N. & Frenkel, D. Fluid–fluid coexistence in colloidal systems with short-ranged strongly directional attraction. *J. Chem. Phys.* **118,** 9882–9889 (2003).
7. Binks, B. P. Particles as surfactants: similarities and differences. *Curr. Opin. Colloid Interface Sci.* **7,** 21–41 (2002).
8. Chen, T., Chen, G., Xing, S., Wu, T. & Chen, H. Scalable routes to Janus Au-SiO_2 and ternary Ag-Au-SiO_2 nanoparticles. *Chem. Mater.* **22,** 3826–3828 (2010).
9. Wang, Y. *et al.* Colloids with valence and specific directional bonding. *Nature* **491,** 51–55 (2012).
10. Pawar, A. B. & Kretzschmar, I. Fabrication, assembly, and application of patchy particles. *Macromol. Rapid Commun.* **31,** 150–168 (2010).
11. Sacanna, S. & Pine, D. J. Shape-anisotropic colloids: building blocks for complex assemblies. *Curr. Opin. Colloid Interf. Sci*. **16,** 96–105 (2011).
12. Nie, Z. H., Petukhova, A. & Kumacheva, E. Properties and emerging applications of self-assembled structures of inorganic nanoparticles. *Nat. Nanotechnol.* **5,** 15–25 (2010).
13. Lattuada, M. & Hatton, T. A. Synthesis, properties and applications of Janus nanoparticles. *Nano Today* **6,** 286–308 (2011).
14. Andala, D. M., Shin, S. H. R., Lee, H. Y. & Bishop, K. J. M. Templated synthesis of amphiphilic nanoparticles at the liquid-liquid interface. *ACS Nano* **6,** 1044–1050 (2012).
15. Vilain, C., Goettmann, F., Moores, A., Le Floch, P. & Sanchez, C. Study of metal nanoparticles stabilised by mixed ligand shell: a striking blue shift of the surface-plasmon band evidencing the formation of Janus nanoparticles. *J. Mater. Chem.* **17,** 3509–3514 (2007).
16. Jackson, A. M., Myerson, J. W. & Stellacci, F. Spontaneous assembly of subnanometre ordered domains in the ligand shell of monolayer-protected nanoparticles. *Nat. Mater.* **3,** 330–336 (2004).
17. Bao, C. *et al.* Effect of molecular weight on lateral microphase separation of mixed homopolymer brushes grafted on silica particles. *Macromolecules* **47,** 6824–6835 (2014).
18. Williams, D. R. M. Grafted polymers in bad solvents: octopus surface micelles. *J. Phys. II* **3,** 1313–1318 (1993).
19. Zhulina, E. B., Birshtein, T. M., Priamitsyn, V. A. & Klushin, L. I. Inhomogeneous structure of collapsed polymer brushes under deformation. *Macromolecules* **28,** 8612–8620 (1995).
20. Koutsos, V., van der Vegte, E. W., Pelletier, E., Stamouli, A. & Hadziioannou, G. Structure of chemically end-grafted polymer chains studied by scanning force microscopy in bad-solvent conditions. *Macromolecules* **30,** 4719–4726 (1997).
21. Choi, B. C., Choi, S. & Leckband, D. E. Poly(*N*-isopropyl acrylamide) brush topography: dependence on grafting conditions and temperature. *Langmuir* **29,** 5841–5850 (2013).
22. Erathodiyil, N. & Ying, J. Y. Functionalization of inorganic nanoparticles for bioimaging applications. *Acc. Chem. Res.* **44,** 925–935 (2011).
23. Minelli, C., Lowe, S. B. & Stevens, M. M. Engineering nanocomposite materials for cancer therapy. *Small* **6,** 2336–2357 (2010).
24. Saha, K., Agasti, S. S., Kim, C., Li, X. & Rotello, V. M. Gold nanoparticles in chemical and biological sensing. *Chem. Rev.* **112,** 2739–2779 (2012).
25. Wolf, B. A. & Willms, M. M. Measured and calculated solubility of polymers in mixed-solvents-co-non-solvency. *Makromol. Chem.* **179,** 2265–2277 (1978).
26. Bürgi, T. Properties of the gold–sulphur interface: from self-assembled monolayers to clusters. *Nanoscale* **7,** 15553–15567 (2015).
27. Rubinstein, M. & Colby, R. H. *Polymer Physics* 53 (Oxford Univ. Press, 2003).
28. Zohar, N., Chuntonov, L. & Haran, G. The simplest plasmonic molecules: metal nanoparticle dimers and trimers. *J. Photochem. Photobiol. C* **21,** 26–39 (2014).
29. Choueiri, R., Klinkova, A., Thérien-Aubin, H., Rubinstein, M. & Kumacheva, E. Structural transitions in nanoparticle assemblies governed by competing nanoscale forces. *J. Am. Chem. Soc.* **135,** 10262–10265 (2013).
30. Klinkova, A. *et al.* Structural and optical properties of self-assembled chains of plasmonic nanocubes. *Nano Lett.* **14,** 6314–6321 (2014).
31. Lukach, A., Liu, K., Thérien-Aubin, H. & Kumacheva, E. Controlling the degree of polymerization, bond lengths and bond angles of plasmonic polymers. *J. Am. Chem. Soc.* **134,** 18853–18859 (2012).
32. Khabibullin, A., Mastan, E., Matyjaszewski, K. & Zhu, S. Surface-initiated atom transfer radical polymerization. *Adv. Polym. Sci.* **270,** 29–76 (2015).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank the Connaught Foundation and the National Science and Engineering Research Council of Canada (Discovery and Canada Research Chair programmes) for financial support of this work. E.B.Z. acknowledges partial support from the Russian Foundation for Basic Research (grant 14-03-00372a) and from the Government of Russian Federation (grant 074-U01). M.R. acknowledges financial support from the National Science Foundation (grants DMR-1309892, DMR-1436201 and DMR-1121107), the National Institutes of Health (grants P01-HL108808 and 1UH2HL123645), and the Cystic Fibrosis Foundation. Research was in part carried out at the Center for Functional Nanomaterials, Brookhaven National Laboratory supported by the US Department of Energy, Office of Basic Energy Sciences (under contract number DE-SC0012704). R.M.C. acknowledges the Natural Sciences and Engineering Research Council of Canada for a PGS-D scholarship. A.K. acknowledges an Ontario Trillium Scholarship. We thank M. Michaelis and S. Lin for initiating preliminary experiments and I. Gourevich for help with imaging experiments.

Author Contributions E.K. and R.M.C. proposed the approach and designed the experiments for nanoparticle surface patterning. R.M.C. synthesized and surface-patterned polystyrene-coated gold nanospheres, nanodumbbells and nanorods, and conducted self-assembly experiments of patchy nanospheres in the presence of excess polymer. E.G. synthesized and surface-patterned polystyrene-capped gold nanospheres and silver nanocubes, as well as poly (*N*-vinyl carbazole)-functionalized nanospheres. E.G. and E.M.L. determined polystyrene grafting density on nanosphere surfaces. H.T.-A. synthesized polystyrene-*co*-PI and poly(4-vinyl pyridine) and surface-patterned nanosphere surfaces. R.M.C., E.G. and H.T.-A. statistically characterized the morphology of patchy nanospheres. A.K. synthesized triangular silver nanoprisms and nanocubes and conducted experiments on nanocube self-assembly. A.Q.-F. surface-patterned gold nanocubes and conducted self-assembly of patchy nanospheres. L.H. performed tomography experiments and H.L.X. and O.G. interpreted the data. E.B.Z. and M.R. developed the theoretical model describing polymer segregation on the nanosphere surface.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to E.K. (ekumache@chem.utoronto.ca).

Reviewer Information *Nature* thanks I. Kretzschmar, D. Williams and the other anonymous reviewer(s) for their contribution to the peer review of this work.

doi:10.1038/nature19786

Cobalt carbide nanoprisms for direct production of lower olefins from syngas

Liangshu Zhong[1], Fei Yu[1,2], Yunlei An[1,2], Yonghui Zhao[1], Yuhan Sun[1,3], Zhengjia Li[1,4], Tiejun Lin[1], Yanjun Lin[1], Xingzhen Qi[1], Yuanyuan Dai[1,5], Lin Gu[6], Jinsong Hu[5,7], Shifeng Jin[6], Qun Shen[1] & Hui Wang[1]

Lower olefins—generally referring to ethylene, propylene and butylene—are basic carbon-based building blocks that are widely used in the chemical industry, and are traditionally produced through thermal or catalytic cracking of a range of hydrocarbon feedstocks, such as naphtha, gas oil, condensates and light alkanes[1,2]. With the rapid depletion of the limited petroleum reserves that serve as the source of these hydrocarbons, there is an urgent need for processes that can produce lower olefins from alternative feedstocks[3–9]. The 'Fischer–Tropsch to olefins' (FTO) process has long offered a way of producing lower olefins directly from syngas—a mixture of hydrogen and carbon monoxide that is readily derived from coal, biomass and natural gas[3–7]. But the hydrocarbons obtained with the FTO process typically follow the so-called Anderson–Schulz–Flory distribution, which is characterized by a maximum C_2–C_4 hydrocarbon fraction of about 56.7 per cent and an undesired methane fraction of about 29.2 per cent (refs 1, 10–12). Here we show that, under mild reaction conditions, cobalt carbide quadrangular nanoprisms catalyse the FTO conversion of syngas with high selectivity for the production of lower olefins (constituting around 60.8 per cent of the carbon products), while generating little methane (about 5.0 per cent), with the ratio of desired unsaturated hydrocarbons to less valuable saturated hydrocarbons amongst the C_2–C_4 products being as high as 30. Detailed catalyst characterization during the initial reaction stage and theoretical calculations indicate that preferentially exposed {101} and {020} facets play a pivotal role during syngas conversion, in that they favour olefin production and inhibit methane formation, and thereby render cobalt carbide nanoprisms a promising new catalyst system for directly converting syngas into lower olefins.

The catalytic mechanism for the FTO process is very similar to that for typical Fischer–Tropsch synthesis, and Fischer–Tropsch catalysts can be used to produce lower olefins after proper catalyst modification and optimization of reaction conditions[3–7,13–17]. Recently, it was reported that iron-based Fischer–Tropsch catalysts promoted by sulfur and sodium exhibited excellent selective formation of lower olefins (which constituted 61% of the carbon products, that is, 61 C%)[3]. Nonetheless, with the goal of simultaneously achieving high selectivity for the production of lower olefins, low selectivity for methane production, and high stability, it is necessary to develop new FTO catalysts that operate away from the Anderson–Schulz–Flory (ASF) distribution.

Here, we prepared cobalt-manganese composite oxide (CoMn catalyst) and investigated it in the FTO reaction (Table 1). This catalyst, after reaching the steady state, displayed a high selectivity for the production of lower olefins (60.8 C%) and a low methane selectivity (5.0 C%) at a CO conversion of 31.8% under mild reaction conditions (250 °C, 1 bar, H_2/CO ratio of 2). This type of catalyst was even suitable for converting coal- and biomass-derived syngas with low H_2/CO ratios. When the H_2/CO ratio was decreased from 2 to 0.5, methane selectivity decreased to 2.4 C%, and the olefin/paraffin ratio for C_2–C_4 slate increased to 51, while selectivity for the production of lower olefins remained high (45.1 C%). (A high olefin/paraffin ratio is of significance industrially; when the ratio is high enough, the paraffin can be ignored and the product used as if it were pure olefins.) The product distribution deviated greatly from the classical ASF distribution, with the highest selectivity being for propylene (Extended Data Fig. 1). Higher total pressures increased the selectivity for the production of oxygenates and decreased the olefin/paraffin ratio, owing to the enhanced second hydrogenation of olefins (Table 1). Obviously, a lower H_2/CO ratio ($\leq$1) and low reaction pressure ($\leq$5 bar) resulted in greater olefin formation. We also carried out a stability test under optimized reaction conditions from an industrial viewpoint; there was no obvious deactivation over 600 h or more (Extended Data Fig. 2), suggesting promising potential industrial application.

These steady-state results differ from the typical performance of a cobalt-based Fischer–Tropsch catalyst. Even more noteworthy is how the catalytic performance evolved during the reaction (Fig. 1a and Extended Data Table 1). Over the first 20 h, both the CO-conversion rate and the product selectivity changed considerably. At the beginning,

Table 1 | Catalytic performance of the CoMn catalyst at different H_2/CO ratios and reaction pressures

Pressure (bar)	H_2/CO ratio	CO conversion (C%)	CO_2 selectivity (C%)	Product selectivity (C%, CO_2-free)					Olefin/paraffin ratio			
				CH_4	$C_{2-4}^{=}$	C_{2-4}^{o}	C_{5+}	Oxy.	C_2	C_3	C_4	C_{2-4}
1	2	31.8	47.3	5.0	60.8	2.0	31.4	0.8	19	41	29	30
1	1	11.5	48.0	3.7	50.0	1.3	43.5	1.5	27	53	36	40
1	0.5	6.3	48.3	2.4	45.1	0.9	48.7	2.0	35	68	46	51
3	0.5	14.3	48.4	4.2	44.3	2.9	38.6	10.0	7	26	20	15
5	0.5	23.6	48.0	3.7	41.2	4.7	35.7	14.7	4	13	10	8
10	0.5	28.6	46.6	4.6	31.9	5.7	39.1	18.6	2	8	6	5

Reaction conditions: 250 °C, 2,000 ml h^{-1} g_{cat}^{-1} (where g_{cat} denotes grams of catalyst). Oxy., oxygenates.

[1]CAS Key Laboratory of Low-Carbon Conversion Science and Engineering, Shanghai Advanced Research Institute, Chinese Academy of Sciences, Shanghai 201203, PR China. [2]School of Environmental and Chemical Engineering, Shanghai University, Shanghai 200444, PR China. [3]School of Physical Science and Technology, ShanghaiTech University, Shanghai 201203, PR China. [4]School of Chemistry and Molecular Engineering, East China Normal University, Shanghai 200062, PR China. [5]University of the Chinese Academy of Sciences, Beijing 100049, PR China. [6]Beijing National Laboratory for Condensed Matter Physics, Institute of Physics, Chinese Academy of Sciences, Beijing 100190, PR China. [7]Beijing National Laboratory for Molecular Sciences, Key Laboratory of Molecular Nanostructure and Nanotechnology, Institute of Chemistry, Chinese Academy of Sciences, Beijing 100190, PR China.

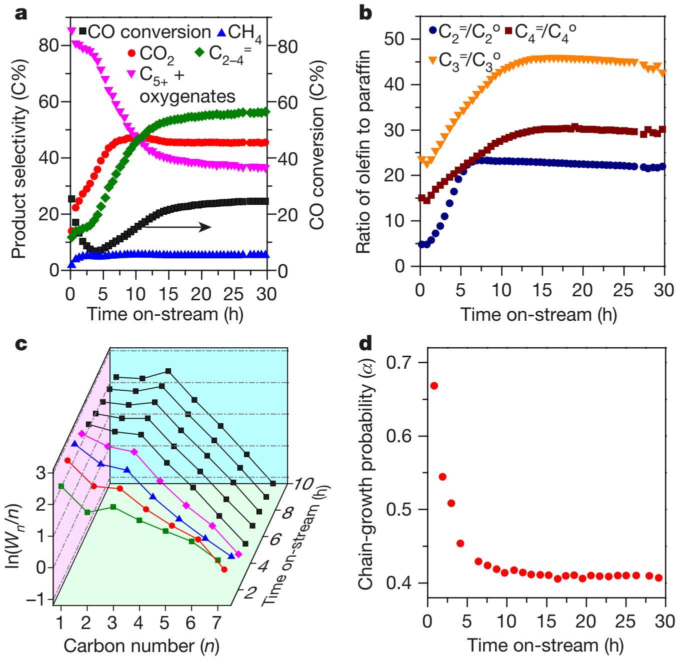


Figure 1 | Catalytic performance of the CoMn catalyst in the initial stages of the reaction. a, CO conversion and product selectivity as a function of time on-stream. Products are CH_4, CO_2, $C_{2-4}^{=}$, and C_{5+} + oxygenates. **b**, Ratio of olefin to paraffin as a function of time on-stream. **c**, Product plots ($\ln(W_n/n)$ and n) versus time on-stream. W_n is the fraction, by weight, of a product with n carbon atoms. **d**, Probability of the growth of carbon chains (α) as a function of time on-stream, obtained by fitting the results generated for chains of three to seven carbons using the ASF model. Reaction conditions: 250 °C, 1 bar, 2,000 ml h^{-1} g_{cat}^{-1}, $H_2/CO = 2$.

after the syngas feed was introduced, high activity and high C_{5+} selectivity were achieved, while the selectivity for light hydrocarbons and CO_2 was rather low. Over the next 4 h, CO conversion dramatically decreased and then reached a minimum value (6.8% at 4 h). During this stage (0–4 h), the selectivity for CO_2 increased gradually. Regarding the distribution of the hydrocarbon products, methane production first increased and then remained stable at around 5 C%; meanwhile, the production of lower olefins increased continuously at the expense of C_{5+} production. In the following stage (4–15 h), the catalytic activity increased markedly and the product distribution shifted greatly from heavy hydrocarbons to lower olefins. After 15 h, the catalytic performance seemed to be stable, and no obvious change was observed afterwards. CO conversion stabilized at about 24% and CO_2 selectivity remained at about 45 C%. Amongst the hydrocarbon products, the selectivity for lower olefins and methane remained at about 55 C% and 5 C%, respectively. The olefin/paraffin ratio also increased gradually with time on-stream, and then became stable at 15 h (Fig. 1b). At steady state, the ratios of $C_2^{=}/C_2^{o}$, $C_3^{=}/C_3^{o}$ and $C_4^{=}/C_4^{o}$ were 22, 45 and 30, respectively (where $C_n^{=}$ is an olefin product with n carbons, and C_n^{o} is a paraffin product with n carbons).

We plot the distribution of hydrocarbon products as a function of time on-stream in Fig. 1c. The $\ln(W_n/n)$ value of C_1 was higher than that of C_2 at the beginning (W_n is the fraction, by weight, of a carbon product with n carbon atoms). However, the $\ln(W_n/n)$ values of C_1 and C_2 were similar after 10 h. The $\ln(W_n/n)$ value of C_3 was the highest, consistent with the highest selectivity for C_3 product. Generally, the $\ln(W_n/n)$ value of C_1 showed a substantial deviation from the typical ASF distribution and was much lower than the modelled value. Figure 1d shows the variation in the probability of chain growth (α) for hydrocarbons as a function of time on-stream; this probability decreased rapidly within the first 4 h and came to a steady value gradually, corresponding to the variation in hydrocarbon selectivity.

As the performance of the catalyst studied here is very different from that of the typical cobalt-based Fischer–Tropsch catalyst, we reasoned that the active site might not be metallic cobalt. In order to reveal the nature of the active site that favours the formation of lower olefins, and to elucidate the structure–performance relationship, we used X-ray diffraction (XRD) to investigate the structure of spent catalyst samples at different stages of the reaction (Extended Data Fig. 3). After the syngas feed was introduced, the metallic cobalt increasingly turned into cobalt carbide (Co_2C) with time on-stream. In the meantime, cobalt and manganese, initially present as a single phase ($Co_xMn_{1-x}O$), segregated into Co_2C and MnO, respectively. Quantitative analysis of the different phases (Co, Co_2C, $Co_xMn_{1-x}O$ and MnO) at different stages of the reaction is shown in Extended Data Fig. 3h. For the freshly reduced sample, the relative contents of metallic Co and $Co_xMn_{1-x}O$ were 3.9% and 96.1%, respectively. After 2 h of reaction, almost all of the metallic Co was transformed into Co_2C; there was also a small amount of MnO. As the reaction proceeded, the amount of Co_2C and MnO increased further while that of $Co_xMn_{1-x}O$ decreased gradually. The existence of metallic Co in the fresh reduced catalyst was responsible for the low CO_2 selectivity, low methane selectivity and high selectivity for long-chain hydrocarbons during the initial reaction stage (before 4 h). The fact that the Co_2C content increased with reaction time suggests that Co_2C might be the active phase for the FTO reaction.

It is often considered that the formation of Co_2C is responsible for the deactivation of cobalt-based Fischer–Tropsch catalysts[18–23]; in these studies, the Co_2C nanoparticles exhibited nanosphere-like morphology. Therefore, we prepared Co_2C nanoparticles with nanosphere-like morphology as described[18], and tested them for syngas conversion (Extended Data Fig. 4). We found relatively low CO-hydrogenation activity and very high selectivity to methane (about 80 C%) for such Co_2C nanospheres, consistent with ref. 18. However, the Co_2C formed from our CoMn catalyst exhibited much higher activity and promising product selectivity during the FTO reaction, suggesting that there could be some essential differences in the form of the Co_2C phase. Representative transmission electron micrograph (TEM) pictures of our as-prepared CoMn catalyst are shown in Fig. 2a–e and Extended Data Fig. 5a–d, and indicate that some nanoparticles presented a quadrangular nanoprism morphology with the shape of a parallelepiped. Further high-resolution (HR)-TEM characterizations and TEM-energy-dispersive X-ray (EDX) analyses demonstrate that such nanoprisms were composed of Co_2C, with specific exposed facets of planar geometry (101), (−101) and (020) (Fig. 2c–e and Extended Data Fig. 5e–h). In addition, sphere-like nanoparticles were composed of MnO or $Co_xMn_{1-x}O$ (Extended Data Fig. 5c, d). The geometric model in Fig. 2f illustrates that the Co_2C nanoprisms had a parallelepiped morphology with a pair of rhomboid faces. We ascribe the pair of rhomboid faces to the facets of (020), and the other rectangles to the facets of (101) and (−101). On the basis of TEM observations at different reaction times (Extended Data Fig. 6), we suggest that the Co_2C nanocrystals were formed and gradually evolved into a nanoparallelepiped shape as the initial reaction proceeded (especially from 0 h to 10 h)—consistent with the change in catalytic performance.

Alkali elements and manganese have a 'promoter' effect, improving the activity and selectivity of Fischer–Tropsch synthesis[6,24–27]. For our CoMn catalyst, prepared by co-precipitation using sodium carbonate, the residual sodium (0.33 wt%, as judged by inductively coupled plasma (ICP) mass spectrometry; Extended Data Fig. 7) and manganese might have a large effect on the catalyst structure and catalytic performance. On the basis of a control study (Extended Data Figs 7, 8), we suggest that the residual sodium enhances the formation of Co_2C, while the manganese contributes to the formation of nanoprisms, via the $Co_xMn_{1-x}O$ precursor. As the catalytic performance of our Co_2C nanoprisms was substantially different from that of reported sodium (or manganese)-promoted, iron (or cobalt)-based Fischer–Tropsch catalysts, we suggest that Co_2C nanoprisms with exposed facets of {101} and {020} geometry represent a new active phase for syngas conversion.

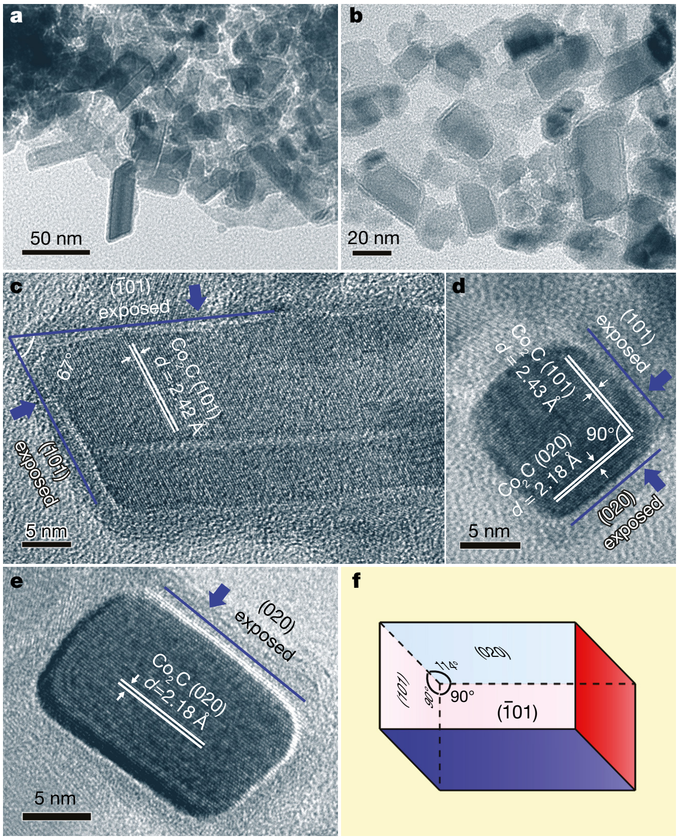


Figure 2 | TEM images of the CoMn catalysts after reaching steady state. **a**, **b**, Low-resolution TEM images. **c**–**e**, High-resolution images of Co_2C nanoprisms with exposed facets of (101), (−101) and (020). *d*, distance (length) of the lattice fringes. **f**, The Co_2C nanoprism has a parallelepiped shape, with four rectangular faces and two rhomboid faces.

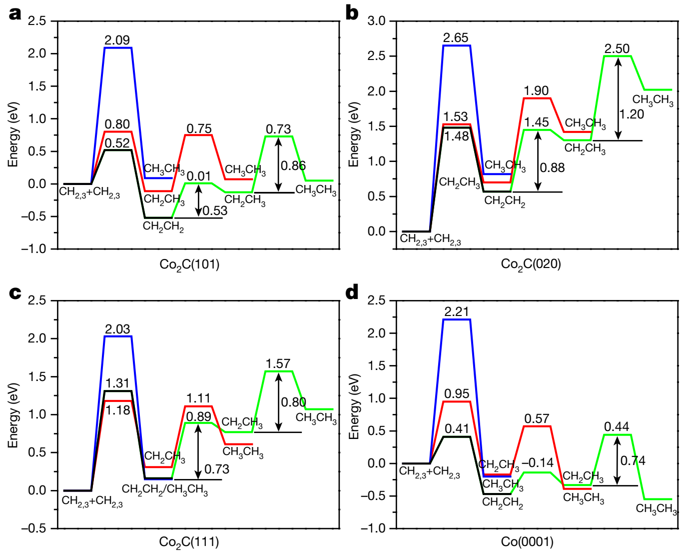


Figure 3 | Energy profiles for pathways that lead to the formation of CH_2CH_2 and CH_3CH_3, on different surfaces of Co_2C and Co. **a**, The $Co_2C(101)$ surface. **b**, The $Co_2C(020)$ surface. **c**, The $Co_2C(111)$ surface. **d**, The Co(0001) surface. The intermediate state of $CH_x + CH_x$ (where $x = 2$ or 3) on the surfaces are chosen as the zero point for all the energy profiles. The path in black is $CH_2 + CH_2 \rightarrow CH_2CH_2$; the path in green is $CH_2 + CH_2 + 2H \rightarrow CH_2CH_2 + 2H \rightarrow CH_2CH_3 + H \rightarrow CH_3CH_3$; the path in red is $CH_2 + CH_3 + H \rightarrow CH_2CH_3 + H \rightarrow CH_3CH_3$; and the path in blue is $CH_3 + CH_3 \rightarrow CH_3CH_3$.

To shed further light on the origin of the promising catalytic performance of the Co_2C nanoprisms, we resorted to density functional theory (DFT) calculations of the formation of lower olefins (CH_2CH_2 as an example) and of methane on $Co_2C(101)$, $Co_2C(020)$, $Co_2C(111)$ and Co(0001) surfaces (Extended Data Fig. 9a). Taking $CH_x + CH_x + H$ (where $x = 2$ or 3) as starting points, we examined three possible paths to CH_3CH_3 formation on the $Co_2C(101)$, $Co_2C(020)$, $Co_2C(111)$ and Co(0001) surfaces, and two paths to CH_2CH_2 formation. From the energy profiles in Fig. 3, we can see that the major chain-growth pathways vary from one surface to the other and are associated with different kinetic barriers and reaction thermodynamics. For $Co_2C(101)$, $Co_2C(020)$, $Co_2C(111)$ and Co(0001), the calculated barriers for CH_2CH_2 formation via coupling of CH_2 are 0.52 eV, 1.48 eV, 1.31 eV and 0.41 eV, respectively. The overall reaction energies are −0.52 eV, 0.57 eV, 0.16 eV and −0.47 eV, respectively. The effective barriers (ΔE_{eff}) for the formation of CH_3CH_3 from the zero point ($CH_2 + CH_2$) are 0.73 eV, 2.50 eV, 1.57 eV and 0.44 eV, respectively. Figure 3a–c also shows that CH_2CH_2 is the most stable intermediate on the $Co_2C(101)$, $Co_2C(020)$ and $Co_2C(111)$ surfaces during the overall pathways; CH_3CH_3 is the most stable intermediate on Co(0001) (Fig. 3d). We also investigated the hydrogenation pathway from CH_x ($x = 2$ or 3) species to methane. The effective barriers for methane formation from CH_2 and H on $Co_2C(101)$, $Co_2C(020)$, $Co_2C(111)$ and Co(0001) surfaces are 0.89 eV, 1.87 eV, 1.39 eV and 0.82 eV, respectively (Extended Data Fig. 9b). Methane formation is exothermic only on the Co(0001) surface.

On the basis of these DFT calculations, we conclude that only on the $Co_2C(101)$ surface does CH_2CH_2 remain the most stable species, from both a thermodynamic and a kinetic point of view. Meanwhile, it is difficult to form methane on both $Co_2C(101)$ and $Co_2C(020)$. So, these theoretical calculations are consistent with the experimental finding that Co_2C nanoprisms formed from the initial catalyst preferentially expose facets of {101} or {020}, which exhibit high selectivity to olefins and low selectivity to methane.

In conclusion, we have demonstrated that Co_2C nanoprisms with exposed {101} and {020} facets exhibit high selectivity towards the formation of lower olefins and low selectivity towards methane production under mild reaction conditions. The olefin/paraffin ratio for C_{2-4} slate is as high as 30, and the product distribution deviates markedly from the classical ASF distribution. As the facet geometry of Co_2C clearly has a very strong effect on syngas conversion, we anticipate that morphological control of the Co_2C nanostructure, with preferential exposure of {101} and {020}, should guide the design of the next generation of highly efficient FTO catalysts.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 15 April; accepted 15 August 2016.

1. Torres Galvis, H. M. & de Jong, K. P. Catalysts for production of lower olefins from synthesis gas: a review. *ACS Catal.* **3,** 2130–2149 (2013).
2. Corma, A., Melo, F. V., Sauvanaud, L. & Ortega, F. Light cracked naphtha processing: controlling chemistry for maximum propylene production. *Catal. Today* **107–108,** 699–706 (2005).
3. Torres Galvis, H. M. *et al.* Supported iron nanoparticles as catalysts for sustainable production of lower olefins. *Science* **335,** 835–838 (2012).
4. Torres Galvis, H. M. *et al.* Iron particle size effects for direct production of lower olefins from synthesis gas. *J. Am. Chem. Soc.* **134,** 16207–16215 (2012).
5. Koeken, A. C. J., Torres Galvis, H. M., Davidian, T., Ruitenbeek, M. & de Jong, K. P. Suppression of carbon deposition in the iron-catalyzed production of lower olefins from synthesis gas. *Angew. Chem. Int. Ed.* **51,** 7190–7193 (2012).
6. Torres Galvis, H. M. *et al.* Effects of sodium and sulfur on catalytic performance of supported iron catalysts for the Fischer–Tropsch synthesis of lower olefins. *J. Catal.* **303,** 22–30 (2013).
7. Santos, V. P. *et al.* Metal organic framework-mediated synthesis of highly active and stable Fischer–Tropsch catalysts. *Nature Commun.* **6,** 6451 (2015).
8. Jiao, F. *et al.* Selective conversion of syngas to light olefins. *Science* **351,** 1065–1068 (2016).
9. Cheng, K. *et al.* Direct and highly selective conversion of synthesis gas to lower olefins: design of a bifunctional catalyst combining methanol synthesis and carbon-carbon coupling. *Angew. Chem. Int. Ed.* **55,** 4725–4728 (2016).

10. de Smit, E. & Weckhuysen, B. M. The renaissance of iron-based Fischer–Tropsch synthesis: on the multifaceted catalyst deactivation behaviour. *Chem. Soc. Rev.* **37,** 2758–2781 (2008).
11. Zhang, Q., Kang, J. & Wang, Y. Development of novel catalysts for Fischer–Tropsch synthesis: tuning the product selectivity. *ChemCatChem* **2,** 1030–1058 (2010).
12. van der Laan, G. P. & Beenackers, A. A. C. M. Kinetics and selectivity of the Fischer–Tropsch synthesis: a literature review. *Catal. Rev.* **41,** 255–318 (1999).
13. Lu, J. *et al.* Promotion effects of nitrogen doping into carbon nanotubes on supported iron Fischer–Tropsch catalysts for lower olefins. *ACS Catal*. **4,** 613–621 (2014).
14. Liu, Y., Chen, J., Bao, J. & Zhang, Y. Manganese-modified Fe_3O_4 microsphere catalyst with effective active phase of forming light olefins from syngas. *ACS Catal*. **5,** 3905–3909 (2015).
15. Zhou, X. *et al.* Hierarchical structured α-Al_2O_3 supported S-promoted Fe catalysts for direct conversion of syngas to lower olefins. *Chem. Commun.* **51,** 8853–8856 (2015).
16. Chen, X., Deng, D., Pan, X., Hu, Y. & Bao, X. N-doped graphene as an electron donor of iron catalysts for CO hydrogenation to light olefins. *Chem. Commun.* **51,** 217–220 (2015).
17. Cheng, Y. *et al.* Fischer–Tropsch synthesis to lower olefins over potassium-promoted reduced graphene oxide supported iron catalysts. *ACS Catal.* **6,** 389–399 (2016).
18. Mohandas, J. C. *et al.* Fischer–Tropsch synthesis: characterization and reaction testing of cobalt carbide. *ACS Catal.* **1,** 1581–1588 (2011).
19. Cheng, J. *et al.* Density functional theory study of iron and cobalt carbides for Fischer–Tropsch synthesis. *J. Phys. Chem. C* **114,** 1085–1093 (2010).
20. Pei, Y. P. *et al.* High alcohols synthesis via Fischer–Tropsch reaction at cobalt metal/carbide interface. *ACS Catal.* **5,** 3620–3624 (2015).
21. Karaca, H. *et al.* Structure and catalytic performance of Pt-promoted alumina-supported cobalt catalysts under realistic conditions of Fischer–Tropsch synthesis. *J. Catal.* **277,** 14–26 (2011).
22. Claeys, M. *et al.* In situ magnetometer study on the formation and stability of cobalt carbide in Fischer–Tropsch synthesis. *J. Catal.* **318,** 193–202 (2014).
23. Kwak, G. *et al.* In situ monitoring during the transition of cobalt carbide to metal state and its application as Fischer–Tropsch catalyst in slurry phase. *J. Catal.* **307,** 27–36 (2013).
24. Huo, C. F. *et al.* The mechanism of potassium promoter: enhancing the stability of active surfaces. *Angew. Chem. Int. Ed.* **50,** 7403–7406 (2011).
25. Lillebø, A. H., Patanou, E., Yang, J., Blekkan, E. A. & Holmen, A. The effect of alkali and alkaline earth elements on cobalt based Fischer–Tropsch catalysts. *Catal. Today* **215,** 60–66 (2013).
26. Johnson, G. R., Werner, S. & Bell, A. T. An investigation into the effects of Mn promotion on the activity and selectivity of Co/SiO_2 for Fischer–Tropsch synthesis: evidence for enhanced CO adsorption and dissociation. *ACS Catal.* **5,** 5888–5903 (2015).
27. Feltes, T. E. *et al.* Selective adsorption of manganese onto cobalt for optimized $Mn/Co/TiO_2$ Fischer–Tropsch catalysts. *J. Catal.* **270,** 95–102 (2010).

Acknowledgements This work has been supported by the Natural Science Foundation of China (grants 21403278, 21403277, 21573271, 91545112), the Shanghai Municipal Science and Technology Commission, China (grants 15DZ1170500, 14ZR1444600), Shanxi Lu'an Coal Corporation Limited, the Ministry of Science and Technology of China (grant 2016YFA0202802) and the Chinese Academy of Sciences (grant QYZDB-SSW-SLH035, the Youth Innovation Promotion Association of CAS).

Author Contributions L.Z. and Y.S. designed the study, analysed the data and wrote the paper. F.Y. and Y.A. prepared the samples and drafted the manuscript. Y.Z. performed DFT calculations. Z.L. and T.L. studied the promoter effect. Y.L., X.Q. and Y.D. performed catalytic evaluation. L.G., J.H., S.J., Q.S. and H.W. characterized the samples. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to L.Z. (zhongls@sari.ac.cn) or Y.S. (sunyh@sari.ac.cn).

Reviewer Information *Nature* thanks M. Claeys, A. Holmen and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Catalyst preparation. We prepared CoMn catalyst with a spinel structure ($Co_xMn_{3-x}O_4$) by a co-precipitation method. Typically, an appropriate amount of cobalt nitrate ($Co(NO_3)_2 \cdot 6H_2O$; Sinopharm Chemical Reagent Co.) and manganese nitrate (50 wt% $Mn(NO_3)_2$, aqueous solution; Sinopharm) were dissolved in deionized water to form a 2 M mixed salt solution (Co/Mn (mol/mol) = 2/1). Meanwhile, sodium carbonate anhydrous (Na_2CO_3; Sinopharm) was dissolved in deionized water, resulting in a 2 M alkali solution as precipitant. These salt and alkali solutions were simultaneously added, dropwise, into a beaker containing 100 ml deionized water under continuous stirring. A constant pH of 8.0 ± 0.1 and temperature of 30 ± 1 °C were maintained during the precipitation process. After ageing for 2 h at 30 °C, the obtained suspension was centrifuged and washed with deionized water seven times, then dried at 100 °C for 10 h. The samples were then calcined in a muffle furnace at 330 °C for 3 h under static air. In order to investigate the effects of sodium and manganese, we also prepared four other catalysts (Co_3O_4, MnO_2, CoMn-A and CoMn-Na) using similar procedures. Specifically, we prepared Co_3O_4 and MnO_2 by the same method (using Na_2CO_3 for precipitation). CoMn-A represented the CoMn catalyst (Co/Mn = 2/1) prepared by precipitation using $(NH_4)_2CO_3$; CoMn-Na represented the CoMn catalyst (Co/Mn = 2/1) prepared by co-precipitation using $(NH_4)_2CO_3$ followed by impregnation with 0.4 wt% of Na by Na_2CO_3.

Catalyst characterization. Catalyst samples for structural analysis were removed from the reactor at different reaction times, after passivation. Specifically, the feeding gas was switched from syngas (or 10% H_2/Ar) to He and heating was stopped at selected reaction times. After the reactor cooled to room temperature (about 20 °C), a flow of 1% (*v*/*v*) O_2/N_2 was introduced to passivate the catalyst for 1 h, and then the catalyst was removed from the reactor and kept in a sealed glass bottle for structural characterization.

X-ray powder diffraction patterns were used for the purpose of phase characterization and Rietveld refinement. The measurements were performed on a Rigaku Ultima IV X-ray powder diffractometer using Cu Kα radiation with a wavelength of 1.54056 Å at 40 kV and 40 mA. Phase characterizations were realized from powder patterns using the PDF4-2015 database (Co_2C, PDF#03-065-1457; MnO, PDF#01-075-0625; Co, PDF#01-089-7093; CoO, PDF#01-071-1178). Indexing of the compounds was realized from the powder patterns using the DICVOL91 program[28]. Quantitative Rietveld refinement was performed using the FULLPROF program. Structural data for the Co_2C, MnO and $Co_xMn_{1-x}O$ phases were taken from the Inorganic Crystal Structure Database (ICSD, accession numbers 16895, 643195 and 9865; https://icsd.fiz-karlsruhe.de/search/index.xhtml). Lattice parameters for Co_2C and MnO were predetermined in a separate refinement for the 150-hour sample and kept fixed during the refinements. For all specimens, the scale factors, profile shape and broadening parameters, asymmetry and corrections for preferred orientation were refined during the quantitative analysis.

ICP analysis of sodium, cobalt and manganese was performed using an ICP optical emission spectrometer (Perkin Elmer) after sample dissolution according to standard in-house procedures.

TEM measurements were performed on a JEOL JEM 2011 electron microscope with 200 kV accelerating voltage. Samples for TEM were prepared by dispersing the powder in ethanol, followed by ultrasonication. One droplet of the suspension was dripped onto carbon-coated copper grids for measurement. We carried out EDX spectroscopy to locate the elemental distribution of cobalt, manganese and oxygen, using a silicon-drift detector on an ARM-200CF transmission electron microscope (JEOL), operated at 200 keV and equipped with double spherical aberration ('Cs') correctors. The attainable energy resolution of the EDX detector is 130 eV. Spatial drift was corrected with a simultaneous image collector.

Catalytic evaluation. Syngas conversion was carried on at 250 °C in a fixed-bed reactor. Generally, 1.5 g of catalyst, in the form of a mesh of 40–60 units, mixed with 2 ml silica sand of the same size was loaded into a stainless-steel reactor with an inner diameter of 9 mm. Prior to the CO-hydrogenation reaction, the catalyst was reduced *in situ* in 10% H_2/Ar (*v*/*v*) at atmospheric pressure, with a gas-flow rate of 200 ml min^{-1}, at 300 °C for 5 h; the heating ramp was 1 °C min^{-1}. The temperature was then dropped to 250 °C in He (99.999%) flow for 30 min, to purge the residual reduction gas. Subsequently, a mixture of 97 vol% syngas with different H_2/CO (*v*/*v*) ratios (H_2/CO = 2, 1, or 0.5) and 3 vol% N_2 (as an internal standard) was introduced to the reactor as feed gas.

For low-pressure testing of the CO-hydrogenation reaction, a system pressure of 1 bar was adopted. The outlet gas, after passing through the hot trap (120 °C), was immediately analysed online by gas chromatography. In order to obtain detailed information on heavy hydrocarbons and oxygenates, the tail gas was allowed to pass through three absorption bottles in series, using water and toluene as absorbents, for about 24 h, and was then analysed by gas chromatography. The calculated selectivity for oxygenates (mainly aldehydes and alcohols) was about 0.80 C% (free CO_2) when syngas with a H_2/CO ratio of 2 was used, and increased slightly to about 2 C% when the H_2/CO ratio was decreased to 1 or 0.5.

For medium- and high-pressure testing of CO hydrogenation, a system pressure of 3 bars, 5 bars or 10 bars was adopted. The outlet gas passed through a hot trap (120 °C) and a cold trap (−1 °C) to collect aqueous products, liquid oil products and solid wax products. Catalytic activity and product selectivities at steady state were determined by gas-chromatographic analysis of products up to C_{45} (oil products and wax products). Analysis of the products confirmed that they include CO_2, olefins, paraffins and oxygenates. A full analysis method was used to determine the selectivity of formation of various products.

We analysed the H_2, N_2, CO, CH_4, and CO_2 content of the outlet gases by using a carbon molecular sieves column (TDX-1) with a thermal conductivity detector (TCD), using helium as the carrier gas. Hydrocarbons with chains of one to ten (C_1 to C_{10}) carbons were analysed using a KCl-modified alumina capillary column (19095P-K25) with an argon carrier and a hydrogen flame ionization detector (FID). CO conversion was calculated according to an internal standard method, assuming that the amount of N_2 remained constant after the reaction. Liquid products were analysed offline with another three gas chromatographs. Aqueous products were analysed using two Porapak Q columns, furnished with an FID (for recognition of C_1–C_5 oxygenates) or TCD (for recognition of H_2O and MeOH). Liquid oil products were analysed using an HP-1 column with N_2 carrier by FID. The mass balance, carbon balance and oxygen balance calculated in each test was between 97% and 101%.

CO conversion was calculated on a carbon-atom basis, as follows:

$$\text{CO conversion} = \frac{CO_{inlet} - CO_{outlet}}{CO_{inlet}} \times 100\%$$

where CO_{inlet} and CO_{outlet} are moles of CO at the inlet and outlet, respectively.

CO_2 selectivity was calculated according to:

$$CO_2 \text{ selectivity} = \frac{CO_{2\ outlet}}{CO_{inlet} - CO_{outlet}} \times 100\%$$

where $CO_{2\ outlet}$ refers to moles of CO_2 at the outlet.

The selectivity of an individual product (hydrocarbon or oxygenate), C_nH_m, in a CO_2-free reaction was obtained according to:

$$C_nH_m \text{ selectivity} = \frac{nC_nH_{m\ outlet}}{CO_{inlet} - CO_{outlet} - CO_{2\ outlet}} \times 100\%$$

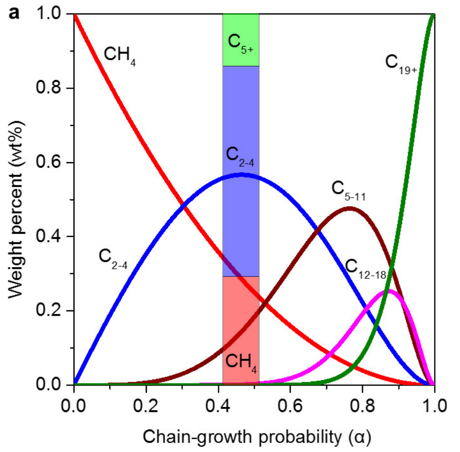
Where $C_nH_{m\ outlet}$ represents moles of individual product (hydrocarbon or oxygenate) at the outlet.

The catalytic performance changed during the initial stage. It was hard to fully analyse all products in the short time period (1 h) at 1 bar, as it would take a longer time for tail-gas treatment with the absorption bottles. In this case, we used a difference method to calculate the selectivity for the production of C_{5+} hydrocarbons (those with five or more carbons) plus oxygenates.

Four repeat experiments were carried out under the same reaction conditions. We found that the catalytic performance of the studied catalyst showed good reproducibility, with both CO conversion and CO_2 selectivity fluctuating within 5%, and the fluctuation in each hydrocarbon selectivity (CO_2-free) being less than 2%.

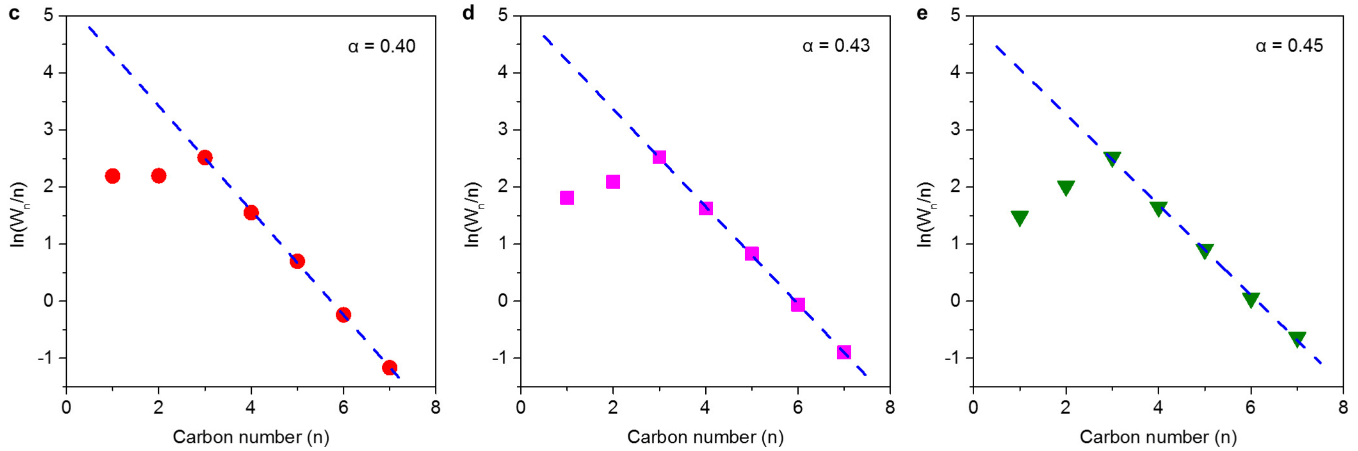
Computations. We performed spin-polarized DFT calculations based on DFT as implemented in the Vienna *ab initio* simulation package (VASP)[29,30] using the projector augmented wave (PAW)[31,32] method. We used Perdew–Burke–Ernzerhof (PBE) functionals[33] to evaluate the non-local exchange-correlation energy, with a planewave energy cutoff of 400 eV. The vacuum region between periodically repeated slabs was 15 Å, and force convergence was 0.02 eV. $p(2 \times 1)$, $p(2 \times 2)$, $p(2 \times 2)$ and $p(3 \times 3)$ unit cells—for $Co_2C(101)$, $Co_2C(020)$, $Co_2C(111)$ and Co(0001) surfaces, respectively—were used in our calculations. A $p(3 \times 3)$ unit cell was applied for the Co(0001) surface. The topmost two equivalent Co layers, including C atoms, are fully relaxed for all of the Co_2C surfaces, while the top three layers of the Co(0001) surface are allowed to relax in all calculations. The adsorbates are also allowed to be fully relaxed throughout. Monkhorst–Pack[34] *k*-point samplings of $(3 \times 4 \times 1)$, $(3 \times 2 \times 1)$, $(3 \times 4 \times 1)$ and $(3 \times 3 \times 1)$ were used in calculations of the $Co_2C(101)$, $Co_2C(020)$, $Co_2C(111)$ and Co(0001) surfaces, respectively. Transition states were located using the force reversed method[35], and a force tolerance of 0.02 eV Å^{-1} was used without zero-point energy correction. Transition states of some of the minimum-energy reaction pathways were located by using the climbing image nudged elastic band (CI-NEB) method[36,37]. Separate adsorption of the species involved in the reactions was adopted for calculation of activation barrier and reaction energy.

28. Boultif, A. & Louër, D. Indexing of powder diffraction patterns for low-symmetry lattices by the successive dichotomy method. *J. Appl. Cryst.* **24,** 987–993 (1991).
29. Kresse, G. & Hafner, J. *Ab initio* molecular dynamics for liquid metals. *Phys. Rev. B* **47,** 558–561 (1993).
30. Kresse, G. & Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B* **54,** 11169–11186 (1996).
31. Blöchl, P. E. Projector augmented-wave method. *Phys. Rev. B* **50,** 17953–17979 (1994).
32. Kresse, G. & Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B* **59,** 1758–1775 (1999).
33. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77,** 3865–3868 (1996).
34. Monkhorst, H. J. & Pack, J. D. Special points for Brillouin-zone integrations. *Phys. Rev. B* **13,** 5188–5192 (1976).
35. Sun, K., Zhao, Y., Su, H. Y. & Li, W. X. Force reversed method for locating transition states. *Theor. Chem. Acc.* **131,** 1118–1127 (2012).
36. Henkelman, G., Uberuaga, B. P. & Jónsson, H. A climbing image nudged elastic band method for finding saddle points and minimum energy paths. *J. Chem. Phys.* **113,** 9901–9904 (2000).
37. Henkelman, G. & Jónsson, H. Improved tangent estimate in the nudged elastic band method for finding minimum energy paths and saddle points. *J. Chem. Phys.* **113,** 9978–9985 (2000).

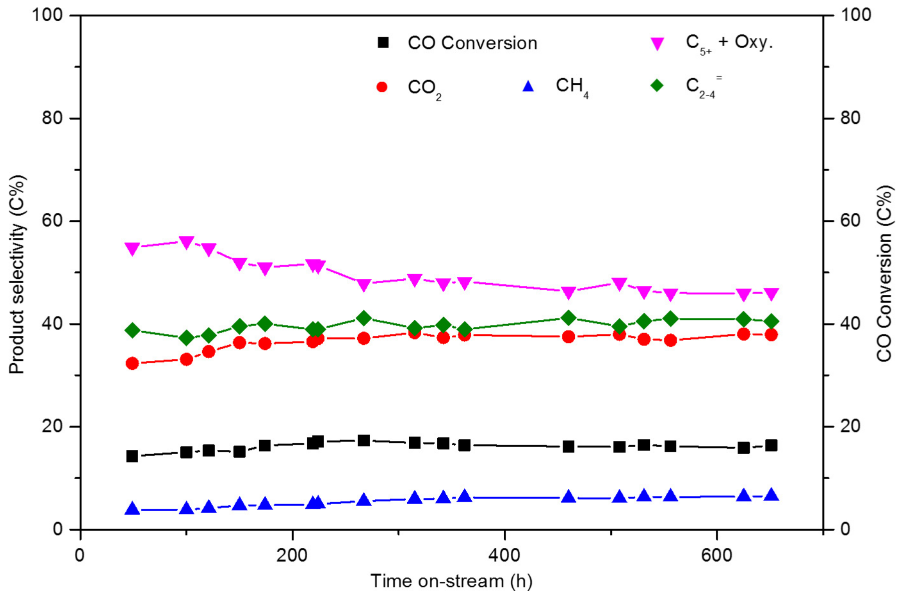


b

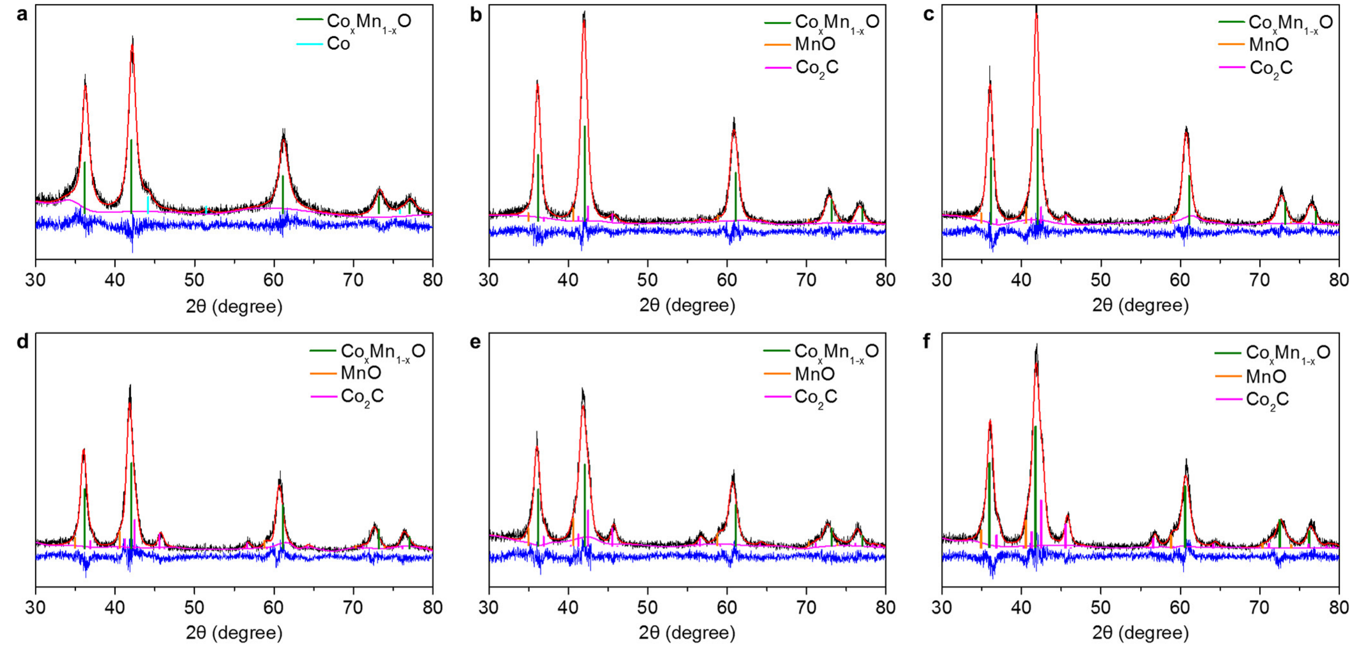
Chain-growth probability α	Hydrocarbon distribution (wt%)					
	CH_4	C_2	C_3	C_4	C_{2-4}	C_{5+}
0.3	49.0	29.4	13.2	5.3	47.9	3.1
0.4	36.0	28.8	17.3	9.2	55.3	8.7
0.46	29.2	26.8	18.5	11.4	56.7	14.1
0.5	25.0	25.0	18.8	12.5	56.3	18.8
0.6	16.0	19.2	17.3	13.8	50.3	33.7
0.7	9.0	12.6	13.2	12.3	38.2	52.8

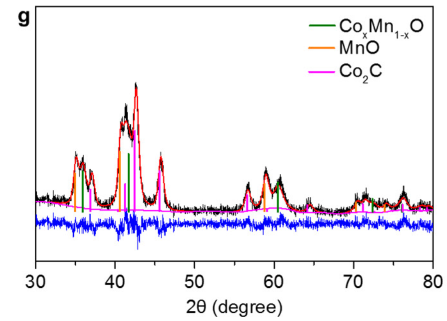


Extended Data Figure 1 | Distribution of products generated by the FTO process from syngas. a, Product distribution according to the Anderson–Schulz–Flory (ASF) model. The vertical bar shows the calculated hydrocarbon distribution for a chain-growth probability (α) of 0.46, which provides the highest proportion of C_2–C_4 hydrocarbons. **b**, Hydrocarbon distribution according to the ASF model for chain-growth probabilities of 0.3, 0.4, 0.46, 0.5, 0.6 and 0.7. **c**–**e**, Typical product plots obtained using our CoMn catalyst (showing $\ln(W_n/n)$ versus n) when the reaction was performed under a H_2/CO ratio of 2 (**c**), 1 (**d**) or 0.5 (**e**), at a temperature of 250 °C and a pressure of 1 bar. The chain-growth probabilities were obtained by fitting the results obtained for C_{3-7} using the ASF model.



Extended Data Figure 2 | Stability test for the CoMn catalyst. Reaction conditions: 250 °C, 3 bar, 6,000 ml h^{-1} g_{cat}^{-1}, H_2/CO = 1. The selectivity of the indicated products remains more or less constant over more than 600 h.

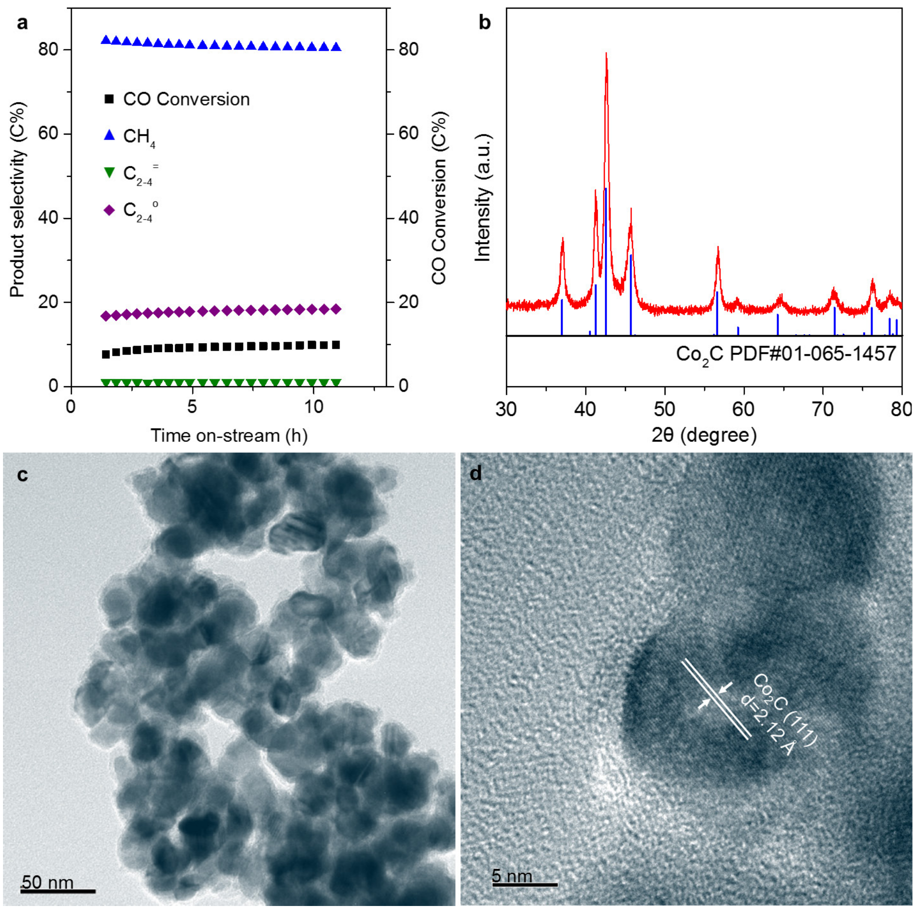




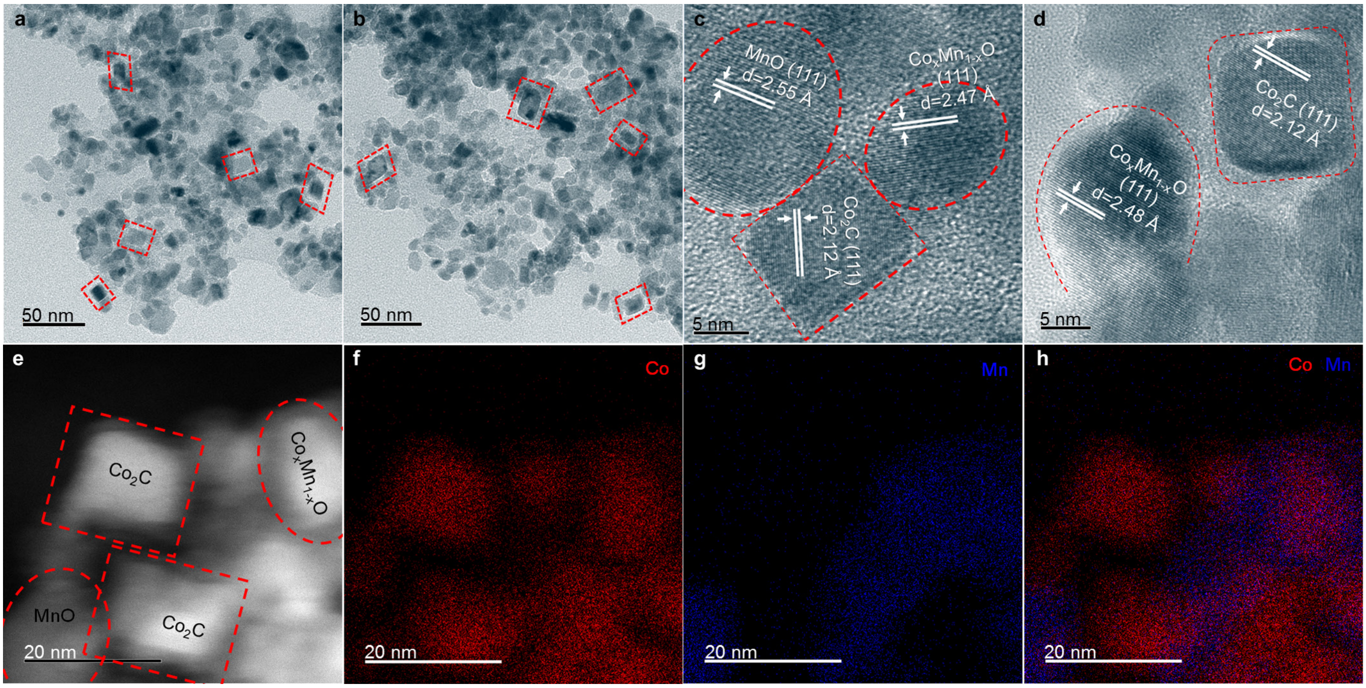
h

Time on-stream (h)	Phase content (wt %)			
	Co-FCC	Co_2C	$Co_xMn_{1-x}O$	MnO
0	3.9	NA	96.1	NA
2	NA	10.3	86.6	3.1
4	NA	13.2	81.8	5.0
10	NA	19.5	71.5	9.0
15	NA	23.1	64.6	12.3
20	NA	26.7	59.9	13.4
150	NA	50.6	21.9	27.5

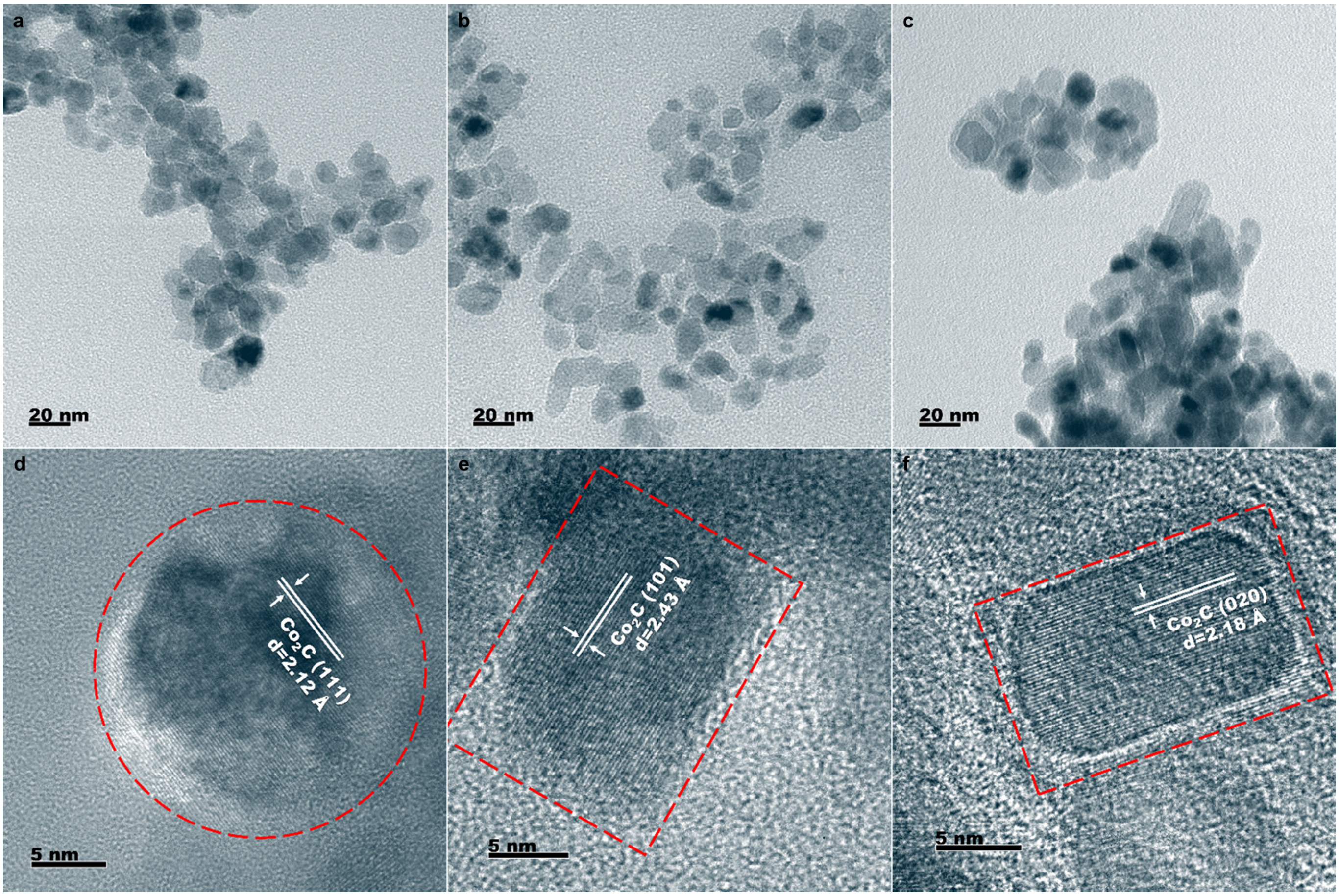
Extended Data Figure 3 | X-ray diffraction analysis of the CoMn catalyst at different times on-stream. a–g, Results from refinement of the X-ray diffraction patterns of catalysts at 0 h (**a**), 2 h (**b**), 4 h (**c**), 10 h (**d**), 15 h (**e**), 20 h (**f**) and 150 h (**g**) ('0 h' refers to the catalyst after reduction). The graphs show the different phases of the catalyst (Co, Co_2C, $Co_xMn_{1-x}O$ and MnO) that are present at different times on-stream. **h**, Quantification of the refinement results on the basis of a full Rietveld analysis. NA, not available.



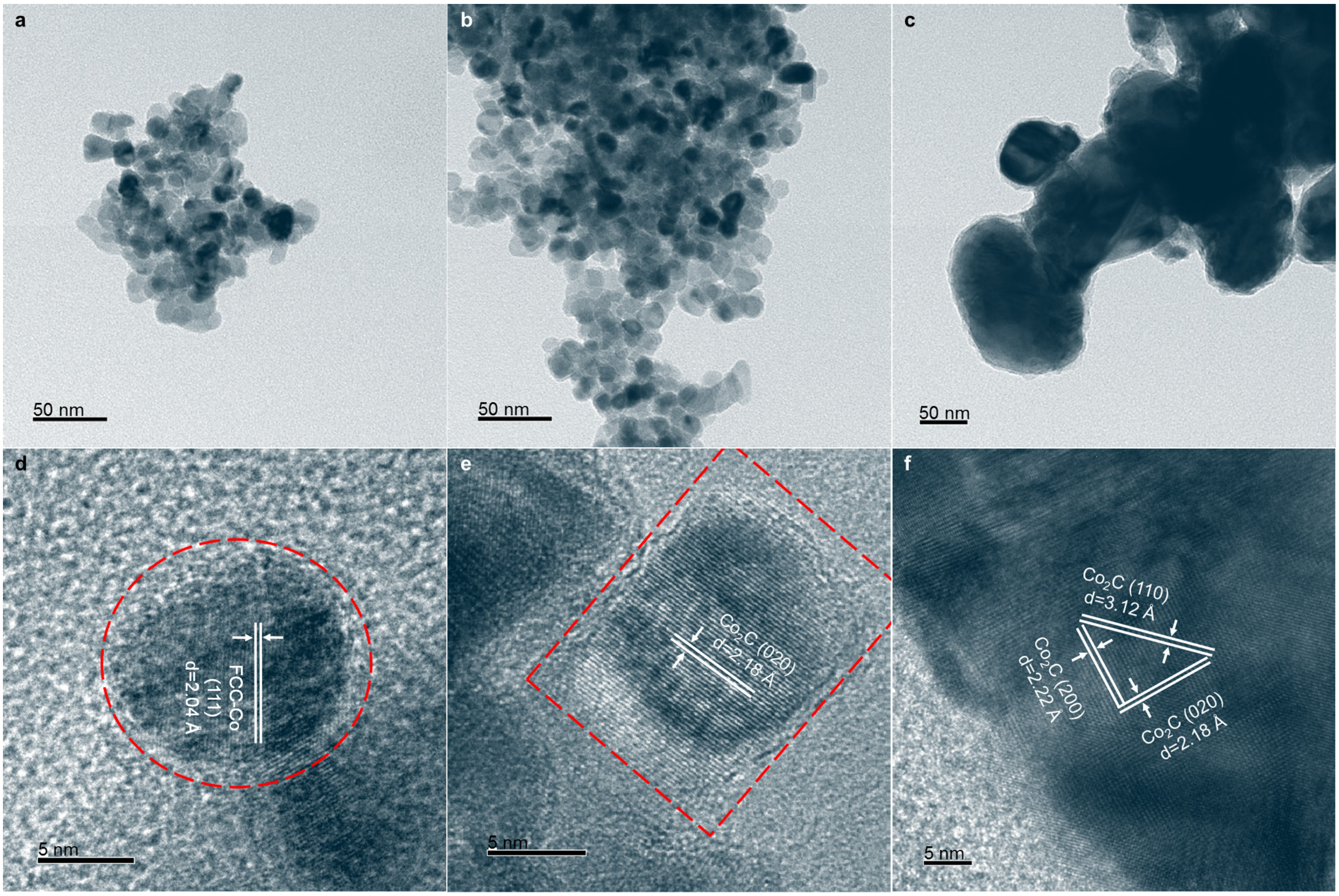
Extended Data Figure 4 | Catalytic performance and structure of Co_2C sphere-like nanoparticles. **a**, Catalytic performance of Co_2C sphere-like nanoparticles with time on-stream. **b**, XRD pattern, **c**, TEM image and **d**, high-resolution TEM image of Co_2C sphere-like nanoparticles. The Co_2C was prepared by carbonizing Co_3O_4 with pure CO at a temperature of 250 °C and at atmospheric pressure for 24 h. The reaction was performed at 250 °C, 1 bar, 2,000 ml h^{-1} g_{cat}^{-1}, $H_2/CO = 2$. The calculated reaction rate for such Co_2C sphere-like nanoparticles was 2.8×10^{-3} mol CO h^{-1} g_{cat}^{-1} for a CO conversion of 9.5%. The calculated reaction rate for the Co_2C catalyst in ref. 18 was 8.9×10^{-4} mol CO h^{-1} g_{cat}^{-1} (CO conversion of 2%, at 20 bar, 220 °C, 3,000 ml h^{-1} g_{cat}^{-1}, $H_2/CO = 2$). The calculated reaction rate for the studied Co_2C nanoprism catalyst was 9.4×10^{-3} mol CO h^{-1} g_{cat}^{-1} (CO conversion of 31.8% at 250 °C, 1 bar, 2,000 ml h^{-1} g_{cat}^{-1}, $H_2/CO = 2$).

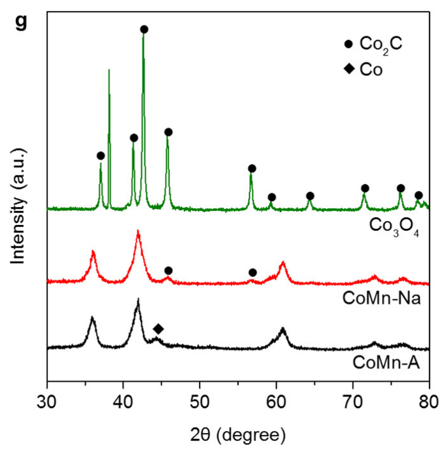


Extended Data Figure 5 | TEM images of the CoMn catalyst after reaching steady state. a–d, Low-resolution images (**a**, **b**) and high-resolution images (**c**, **d**) showing Co_2C nanoprisms (parallelepiped structures) and sphere-like nanoparticles of MnO or $Co_xMn_{1-x}O$. **e**, Scanning TEM image. **f–h**, EDX mapping of Co (**f**), Mn (**g**) and Co plus Mn (**h**). For the nanoprism particles, a higher concentration of Co and a very low concentration of Mn was observed; these results, coupled with the high-resolution TEM images of the lattice fringes, suggest that such nanoprisms are composed of Co_2C. For most of the sphere-like nanoparticles, both Co and Mn were observed, indicating that these particles are composed of CoMn composite oxide ($Co_xMn_{1-x}O_y$). For some of the sphere-like nanoparticles, only Mn was observed and the concentration of Co was very low, suggesting that these particles are composed of MnO.



Extended Data Figure 6 | TEM images of the CoMn catalyst at different reaction times. a–c, Low-resolution TEM images at 2 h (**a**), 10 h (**b**) and 20 h (**c**). **d–f**, Corresponding high-resolution TEM images. For the samples removed at 2 h, it was hard to find Co_2C nanoprisms; the shape of the Co_2C seemed to be irregular. At 10 h, the Co_2C nanostructure seemed to be nanoprism-like, with a parallepiped shape, and the lattice fringes suggested the exposed facet of (101) geometry, although the shape was not perfect. At 20 h, Co_2C nanoprisms were found.





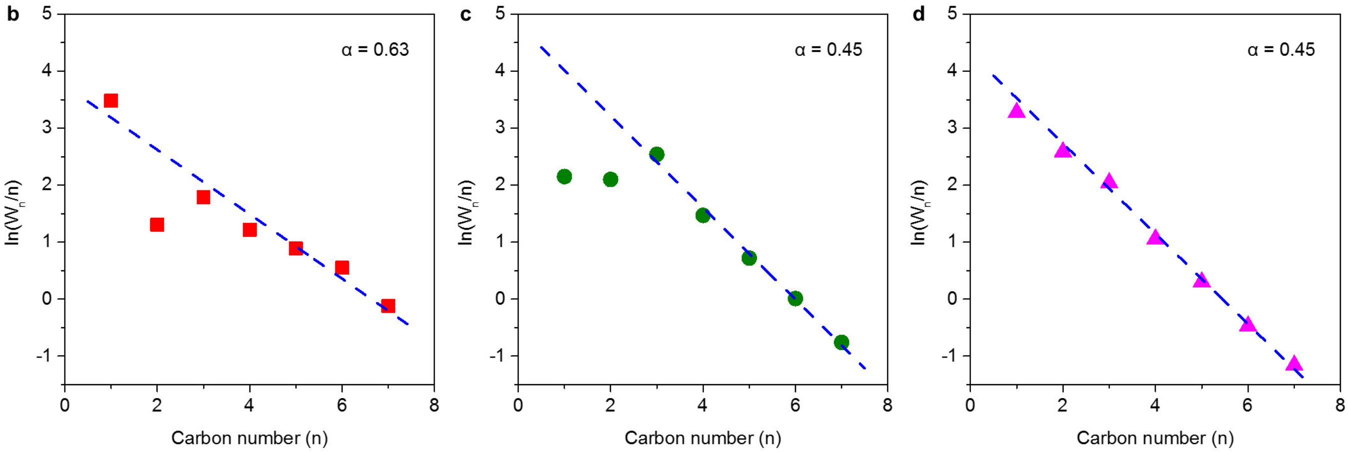
h

Catalysts	Co (wt%)	Mn (wt%)	Na (wt%)
CoMn-A	48.54	22.45	0.006
CoMn-Na	50.18	23.06	0.37
CoMn	48.72	22.17	0.33
Co_3O_4	78.39	0.02	0.16

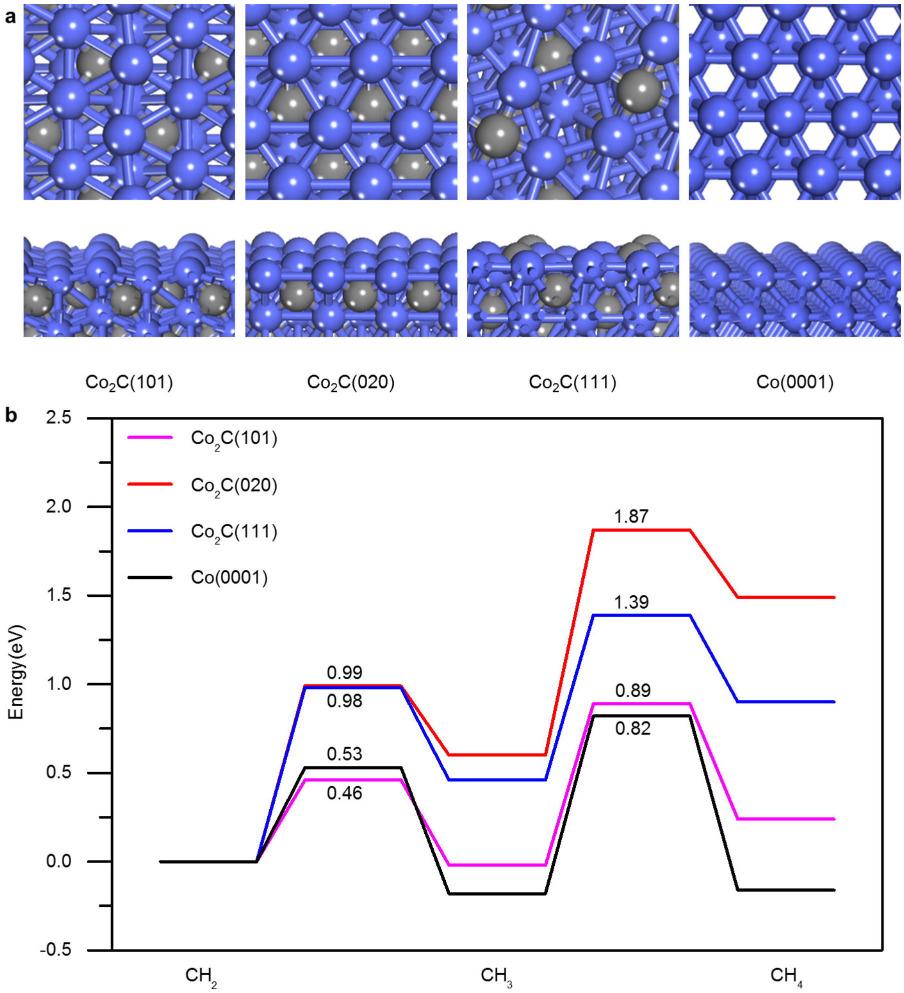
Extended Data Figure 7 | TEM, XRD and ICP analyses of the effects of Na and Mn on different Co-based catalysts. a–c, Low-resolution TEM images of spent CoMn-A (**a**), CoMn-Na (**b**) and Co_3O_4 (**c**). **d–f**, Corresponding high-resolution TEM images. **g**, XRD patterns of spent CoMn-A, CoMn-Na and Co_3O_4 catalysts. **h**, Elemental analysis of the fresh catalysts by ICP mass spectrometry. To produce CoMn-A, CoMn catalyst (Co/Mn = 2/1) was precipitated with $(NH_4)_2CO_3$. To produce CoMn-Na, CoMn catalyst (Co/Mn = 2/1) was precipitated with $(NH_4)_2CO_3$ and impregnated with about 0.4 wt% of Na using Na_2CO_3. CoMn was prepared by precipitating CoMn catalyst (Co/Mn = 2/1) with Na_2CO_3. Co_3O_4 was prepared by precipitation using Na_2CO_3. We found CoMn-A to include sphere-like, face-centred-cubic, metallic Co nanoparticles. We detected Co_2C nanoprisms in CoMn-Na, while Co_3O_4 comprised larger Co_2C sphere-like nanoparticles.

a

Catalysts	CO Conversion (C%)	CO_2 Selectivity (C%)	Product selectivity (C%, CO_2-free)				Olefin/Paraffin			
			CH_4	$C_{2\text{-}4}^{=}$	$C_{2\text{-}4}^{o}$	C_{5+} + Oxy.	C_2	C_3	C_4	$C_{2\text{-}4}$
CoMn-A	46.0	8.3	24.0	22.9	9.6	43.5	0.3	5	4	2
CoMn-Na	20.4	46.5	5.9	54.2	2.3	37.6	12	36	27	24
Co_3O_4	2.8	37.3	23.9	59.4	4.6	12.1	14	14	10	13
MnO_2	—	—	—	—	—	—	—	—	—	—



Extended Data Figure 8 | Catalytic performance and product distribution of different catalysts. a, Performance of different catalysts. **b**–**d**, Product plot (ln(W_n/n) versus n) for CoMn-A (**b**), CoMn-Na (**c**) and Co_3O_4 (**d**) catalysts. CoMn-A, CoMn-Na and Co_3O_4 were prepared as described in Extended Data Fig. 7. MnO_2 was prepared by precipitation using Na_2CO_3. Reaction conditions: 1 bar, 250 °C, 2,000 ml h^{-1} $g_{cat}{}^{-1}$, H_2/CO = 2. There was no detectable CO conversion by MnO_2.



Extended Data Figure 9 | DFT study on different surfaces. **a**, Top views (upper panels) and side views (bottom panels) of different surfaces: from left to right, $Co_2C(101)$, $Co_2C(020)$, $Co_2C(111)$ and Co(0001). Blue, Co atom; grey, C atom. **b**, Energy profiles for CH_4 formation on $Co_2C(101)$, $Co_2C(020)$, $Co_2C(111)$ and Co(0001) surfaces. The intermediate state of $CH_2 + 2H$ is chosen as the zero point for all of the energy profiles.

Extended Data Table 1 | Catalytic performance of CoMn catalyst in the initial stages of the FTO reaction

Time on stream	CO Conversion	CO_2 Selectivity	Product selectivity (C%, CO_2-free)				Olefin/Paraffin			
(h)	(C%)	(C%)	CH_4	$C_{2\text{-}4}^{=}$	$C_{2\text{-}4}^{o}$	C_{5+} + Oxy.	C_2	C_3	C_4	$C_{2\text{-}4}$
0.2	25.3	13.9	1.8	11.8	1.0	85.4	5	24	15	12
1.9	10.3	27.2	4.9	14.8	1.1	79.2	7	25	17	14
4.2	6.8	35.8	4.7	19.7	0.9	74.7	16	32	20	22
10.4	15.5	47.1	5.6	46.7	1.5	46.3	23	44	28	32
15.3	21.7	45.6	5.3	53.3	1.6	39.8	23	45	30	34
20.2	23.6	45.4	5.2	55.2	1.6	37.9	22	45	30	34

Reaction conditions: 250 °C, 1 bar, 2,000 ml h^{-1} g_{cat}^{-1}, H_2/CO = 2.

doi:10.1038/nature19797

Upward revision of global fossil fuel methane emissions based on isotope database

Stefan Schwietzke[1,2], Owen A. Sherwood[3], Lori M. P. Bruhwiler[2], John B. Miller[1,2], Giuseppe Etiope[4,5], Edward J. Dlugokencky[2], Sylvia Englund Michel[3], Victoria A. Arling[1,2], Bruce H. Vaughn[3], James W. C. White[3] & Pieter P. Tans[2]

Methane has the second-largest global radiative forcing impact of anthropogenic greenhouse gases after carbon dioxide, but our understanding of the global atmospheric methane budget is incomplete. The global fossil fuel industry (production and usage of natural gas, oil and coal) is thought to contribute 15 to 22 per cent of methane emissions[1–10] to the total atmospheric methane budget[11]. However, questions remain regarding methane emission trends as a result of fossil fuel industrial activity and the contribution to total methane emissions of sources from the fossil fuel industry and from natural geological seepage[12,13], which are often co-located. Here we re-evaluate the global methane budget and the contribution of the fossil fuel industry to methane emissions based on long-term global methane and methane carbon isotope records. We compile the largest isotopic methane source signature database so far, including fossil fuel, microbial and biomass-burning methane emission sources. We find that total fossil fuel methane emissions (fossil fuel industry plus natural geological seepage) are not increasing over time, but are 60 to 110 per cent greater than current estimates[1–10] owing to large revisions in isotope source signatures. We show that this is consistent with the observed global latitudinal methane gradient. After accounting for natural geological methane seepage[12,13], we find that methane emissions from natural gas, oil and coal production and their usage are 20 to 60 per cent greater than inventories[1,2]. Our findings imply a greater potential for the fossil fuel industry to mitigate anthropogenic climate forcing, but we also find that methane emissions from natural gas as a fraction of production have declined from approximately 8 per cent to approximately 2 per cent over the past three decades.

Our current understanding of the global methane (CH_4) budget stems largely from three-dimensional (3D) inversion studies, which use the trends and gradients of atmospheric CH_4 to infer the spatio-temporal distribution of the total CH_4 source, but atmospheric CH_4 data alone do not include source type information. Source type information comes primarily from bottom-up-derived a priori spatial patterns. 3D inverse models return a posteriori fluxes constrained by the bottom-up source type information for each, especially for large regions that contain a mix of source types, like agriculture, wetlands and fossil fuels. Note that we refer below to CH_4 emissions from total fossil fuels (FF_{tot}) as the sum of CH_4 emissions from fossil fuel industry activities (FF_{ind}) and natural geological seepage (FF_{geo}). To alleviate this problem, some previous 3D inversion[3,4] and box model studies[9,10] have included measurements of atmospheric δ^{13}C-CH_4 (henceforth $\delta^{13}C_{atm}$, where $\delta^{13}C = R_{atm}/R_{std} - 1$ and $R = {}^{13}C/{}^{12}C$) as an additional constraint for better source allocation. Broadly defined source categories—that is, FF_{tot}, microbial, and biomass burning—emit CH_4 with different source signatures[10] (δ^{13}C-CH_4; henceforth $\delta^{13}C_{source}$, including $\delta^{13}C_{FF}$, $\delta^{13}C_{mic}$ and $\delta^{13}C_{BB}$).

The sample sizes of $\delta^{13}C_{source}$ values used in published global CH_4 budgets are either small ($N < 100$, based on cited original measurements) or unknown, uncertainties are rarely applied, and global representativeness is lacking (especially in the tropics and the Southern Hemisphere), but some $\delta^{13}C_{source}$ values have nevertheless taken on canonical status with few references to primary sources (for example, refs 3, 4, 9 and 10; see full list of references in Supplementary Information section 8). We have compiled the most comprehensive $\delta^{13}C_{source}$ database to date (see ref. 14 and Supplementary Information sections 3–5 for complete list of data, metadata and references) including 9,468 $\delta^{13}C_{FF}$, $\delta^{13}C_{mic}$ and $\delta^{13}C_{BB}$ original measurements from the peer-reviewed literature and other publicly available sources to define globally weighted average $\delta^{13}C_{FF}$ (time-dependent), $\delta^{13}C_{mic}$, and $\delta^{13}C_{BB}$ with well defined uncertainties. These data allowed us to revisit the source attribution of global CH_4 emissions since the 1980s by applying an atmospheric box-model to global atmospheric CH_4 and $\delta^{13}C_{atm}$ measurements (and avoiding the use of a priori FF_{tot} and microbial source strengths), thus maximizing the CH_4 and $\delta^{13}C_{atm}$ constraints.

Our box-model applies Monte Carlo techniques to estimate global FF_{tot} and microbial CH_4 emissions and uncertainties as a function of $\delta^{13}C_{source}$, of isotope fractionation during oxidation ($OH + CH_4$), of the uncertainties of both of these values, and of other factors (see Supplementary Table 1). We also estimated FF_{ind} emissions by subtracting FF_{geo} emissions from FF_{tot} emissions. This allowed us to calculate global long-term trends in the Fugitive Emission Rate (FER), which is the fraction of natural gas production lost to the atmosphere through its lifecycle (production, processing, transport and use), and is a critical parameter for evaluating the climatic impact of natural gas as a fuel[9,15].

The $\delta^{13}C_{source}$ weighting procedures are described in detail in Supplementary Information sections 3, 4 and 5, and briefly summarized here. The $\delta^{13}C_{FF}$ samples ($N = 7{,}482$) are representative of natural gas or coal gas from 44 countries, accounting for 82% and 80% of global natural gas and coal production, respectively[16]. Country-specific $\delta^{13}C_{FF}$ distributions for natural gas and coal were weighted by their respective annual production of natural gas, oil (co-produced with natural gas), and coal (Supplementary Information section 5). The time averaged, globally weighted $\delta^{13}C_{FF}$ of -44.0 ± 0.7‰ (one standard deviation, 1 s.d.) is much lighter (about 5‰ lighter) than typical values in previous inverse studies (Fig. 1), although Whiticar *et al.*[17] reported an empirically derived $\delta^{13}C_{natural\ gas/oil}$ value of −44‰ based on unpublished proprietary industry sources. Our relatively light $\delta^{13}C_{natural\ gas/oil}$ value is due to contributions from economically important reservoirs of microbially produced CH_4, and from thermogenic gas originating from low-maturity source rocks, or is associated with oil, typically[18] in the range −45‰ to −55‰. Thermogenic CH_4 formation and microbial methanogenesis also occur in coal beds (both deep and shallow deposits[19]).

Our $\delta^{13}C_{source}$ database[14] contains 1,021 $\delta^{13}C_{mic}$ samples from wetlands, termites, ruminants, rice agriculture and waste/landfill, weighted by their relative contribution to global microbial CH_4

[1]Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, Colorado, USA. [2]NOAA Earth System Research Laboratory, Global Monitoring Division, Boulder, Colorado, USA. [3]Institute of Arctic and Alpine Research, University of Colorado, Boulder, Colorado, USA. [4]Istituto Nazionale di Geofisica e Vulcanologia, Sezione Roma 2, Italy. [5]Faculty of Environmental Science and Engineering, Babes Bolyai University, Cluj-Napoca, Romania.

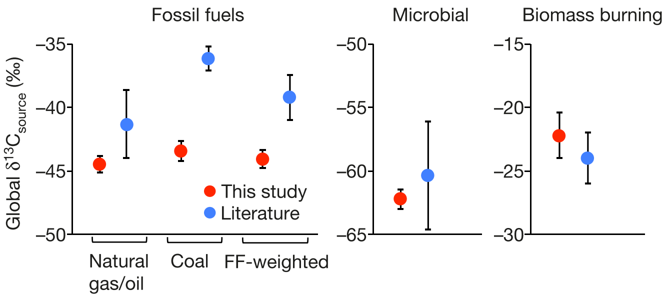


Figure 1 | Comparison of $\delta^{13}C_{FF}$, $\delta^{13}C_{mic}$ and $\delta^{13}C_{BB}$ source signatures from this study (red) and those used in 15 previous top-down studies (blue; see Supplementary Information section 8 for references). Error bars indicate 1 s.d., representing empirical uncertainty in our database (red; $N_{total} = 9{,}468$; see Supplementary Table 1) and variability among mean literature values (blue). $\delta^{13}C_{FF}$ are temporal averages; see Supplementary Fig. 11 for annual $\delta^{13}C_{FF}$ (owing to country-specific FF_{ind} production trends).

emissions based on 12 literature estimates (Supplementary Information section 3). Our $\delta^{13}C_{source}$ database[14] includes 82 direct (plume measurement) and 965 indirect (plant material) $\delta^{13}C_{BB}$ samples from C_3 and C_4 plants, weighted by maps of global vegetation and biomass-burning fluxes (Supplementary Information section 4). Our $\delta^{13}C_{mic}$ and $\delta^{13}C_{BB}$ are within the uncertainty of literature estimates, but have smaller error bars (given the large sample size), and approximately 2‰ differences in mean values (Fig. 1).

Time series of the global microbial and FF_{tot} CH_4 emission distributions are shown in Fig. 2a using 10,000 Monte Carlo box-model repetitions. Estimates of FF_{tot} CH_4 are based only on global average atmospheric CH_4 levels, CH_4 lifetimes, isotope source signatures, isotope fractionation, and soil sink and biomass-burning estimates; that is, they exclude other bottom-up estimates or a priori information. Note that this study focuses on long-term trends, and Fig. 2a presents moving averages. The original model results (Supplementary Fig. 18) include the inter-annual variability of <10% on average, which is partly an artefact of multiple components that our model does not control including CH_4 sink inter-annual variability and the $\delta^{13}C_{BB}$ inter-annual variability depending on the dominant biomass type (C_3 versus C_4) in a given year.

Our updated $\delta^{13}C_{source}$ database causes a substantial upward shift in total fossil fuel CH_4 relative to 'traditional' $\delta^{13}C_{source}$ used in the literature. The approximately 5‰ lighter $\delta^{13}C_{FF}$ alone leads to a FF_{tot} value about 50 Tg CH_4 yr^{-1} greater, and the combination of the approximately 2‰ lighter $\delta^{13}C_{mic}$ and approximately 2‰ heavier $\delta^{13}C_{BB}$ shifts FF_{tot} up by an additional 25 Tg CH_4 yr^{-1} or so (Fig. 2a). Total annual CH_4 emissions increased by about 25 Tg since 2006 (ref. 20, Supplementary Fig. 2), and the microbial source increased by about 45 Tg, partially offset by a FF_{tot} decrease of about 20 Tg to balance the atmospheric CH_4 budget. This growth attribution to microbial sources is mostly driven by changing $\delta^{13}C_{atm}$ (increase stopped around 2000 followed by a 0.2‰ decrease since 2004; see also ref. 21), and a 1.7‰ $\delta^{13}C_{FF}$ increase since 2003 resulting from a redistribution of global FF_{ind} production from countries with different $\delta^{13}C_{FF}$ values. For example, China's rising hard coal production during 1999–2012 increased global coal $\delta^{13}C_{FF}$ by approximately 3.5‰ given China's national average coal $\delta^{13}C_{FF}$ of −36.0‰, and global natural gas $\delta^{13}C_{FF}$ increased by approximately 0.7‰ during 2002–2012 because of rising natural gas

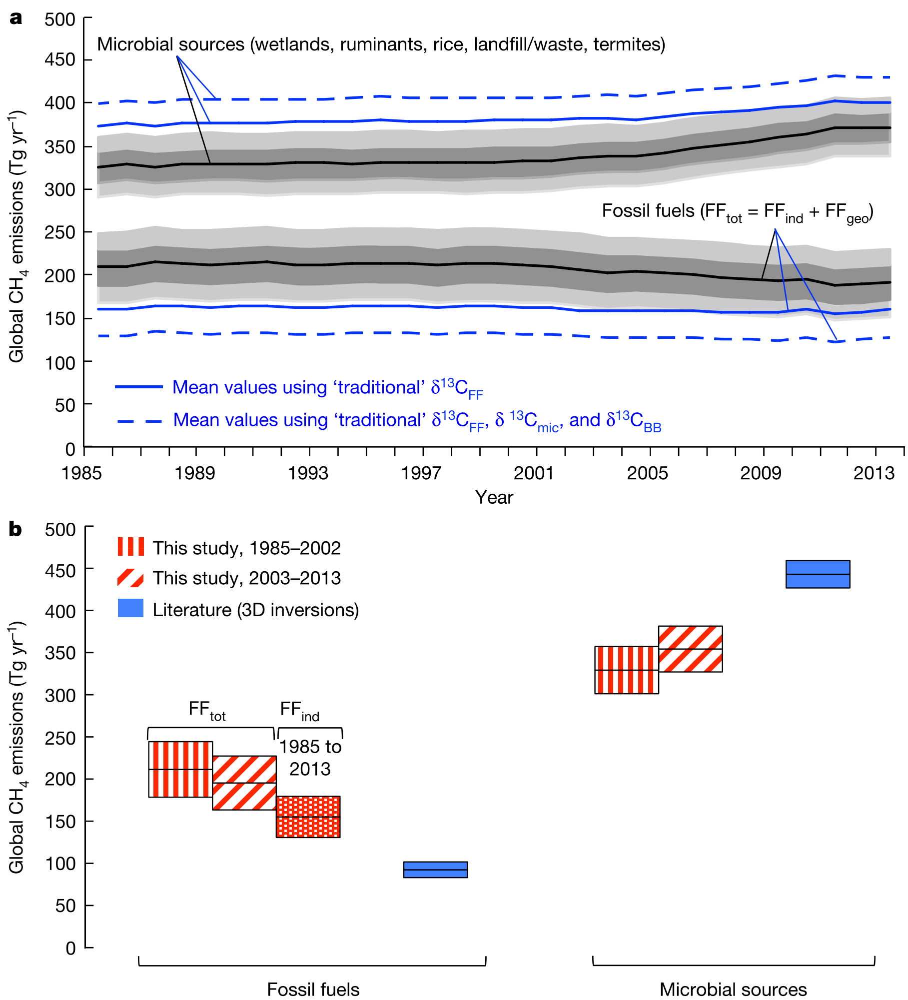


Figure 2 | Fossil fuel and microbial source CH_4 budget terms. a, Long-term trend in global microbial and total FF_{tot} CH_4 emissions from 1985 to 2013. Moving averages are shown; see Supplementary Fig. 18 for original mass balance results including inter-annual variability due to multiple components not accounted for in this model (see text). Mean values are shown in solid black. Dark and light grey bands mark the 25th/75th percentile and the 10th/90th percentiles, respectively. Blue lines assume the mean $\delta^{13}C_{source}$ values from the literature (as in Fig. 1, blue values). See Supplementary Information section 7 for sensitivity analyses. **b**, The box plot compares means and 1 s.d. uncertainties between this study (red) and the recent literature[3–8] 3D inversions (blue). The box-model temporal split marks approximately when the $\delta^{13}C_{atm}$ increase stopped and $\delta^{13}C_{FF}$ decreased; literature periods vary. The literature budget terms were scaled to match this study's mean total CH_4 budget (see Supplementary Information section 8 for individual literature study means). Industrial FF (FF_{ind}) represents total FF (FF_{tot}) minus natural geological seepage (FF_{geo}). Literature $FF_{tot} = FF_{ind}$ because these studies exclude FF_{geo}.

production in multiple countries with relatively heavy natural gas $\delta^{13}C_{FF}$ (details in Supplementary Information section 3).

The FF_{tot} CH_4 decrease is surprising because it marks a period of a dramatic (56%) increase in global coal production, mainly from China[16]. However, not accounting for other factors in our model could produce a downward bias of up to 20 Tg in the FF_{tot} CH_4 trend (Supplementary Fig. 17 for a sensitivity analysis of the model parameters on the budget terms and their trends). For instance, the 0.2‰ $\delta^{13}C_{atm}$ decrease requires decreasing emissions of either FF_{tot} or biomass burning since 2004, and we used a prescribed, temporally constant range of biomass-burning emissions. Constant or decreasing FF_{tot} emissions disagree with emission inventories, suggesting increasing FF_{ind} emissions during this period[1,21]. This may be explained by efficiency improvements such as capturing otherwise vented or flared natural gas, replacement of old equipment, and improved combustion techniques[22], which inventories may not account for completely.

Box-model and literature 3D inverse estimates[3–8] are compared in Fig. 2b. Mean FF_{tot} CH_4 emissions in this study are approximately double previous estimates. The relatively small size of the 3D inverse FF_{tot} CH_4 estimates is partly due to the strong influence of a priori information on source attribution in regions where observations are sparse and multiple sources are co-located. As a result, the inverse FF a posteriori means are on average within 17% of the a priori means[3–8]. Also, these studies are not independent owing to their choice of broadly similar a priori information (these studies use different releases of the EDGAR database[2] for anthropogenic emissions): the a priori mean of five[3–6,8] out of the six studies (Fig. 2b) is only 92 ± 3 Tg CH_4 yr^{-1} (1 s.d.). Note that when $\delta^{13}C_{atm}$ is used in 3D inversions[3,4] (as opposed to box-model studies), the spatio-temporal CH_4 constraint greatly outweighs the information in the relatively sparse $\delta^{13}C_{atm}$ data. For instance, the relative difference in a posteriori FF_{tot} fluxes between including and excluding $\delta^{13}C_{atm}$ data in 3D inversions[3,4] is <7%. Uncertainties in this study (the FF_{tot} 1 s.d. is 32 Tg CH_4 yr^{-1}) are considerably larger than in the inverse studies (9 Tg CH_4 yr^{-1} on average). However, inversion-based a posteriori uncertainties are often derived from relatively narrow a priori source uncertainties, whereas we derive uncertainties directly from sensitivities in the global mass balances of $\delta^{13}C_{atm}$ and CH_4. The only prescribed source in our model is biomass burning, although our prescribed biomass-burning estimates are partly constrained by fire detection using satellites (wild fires can be detected[23], not accounting for fuel biomass burning). These results suggest the route of first obtaining observation-based global total CH_4 source strengths (this study) as inputs to an inversion, which can then use additional spatial information to estimate source allocation at higher resolution.

Until now, most top-down studies have excluded FF_{geo} as an important source of global CH_4 over the past three decades[11], despite bottom-up studies[12,13] as well as top-down analyses using ice-core[24–26] and radiocarbon (^{14}C) data[12] suggesting a FF_{geo} range of 20–76 Tg CH_4 yr^{-1} (4%–14% of the modern global budget; Supplementary Information sections 2 and 6). Constraining these emissions using ice-core CH_4 and $\delta^{13}C$ (henceforth $\delta^{13}C_{ice}$) measurements, and our extensive database of isotope signatures yields FF_{geo} emissions of 51 ± 20 Tg CH_4 yr^{-1} (1 s.d.; Supplementary Information section 6). Subtracting FF_{geo} from FF_{tot} yields mean FF_{ind} emissions of 156 ± 24 Tg CH_4 yr^{-1} (the 1 s.d. uncertainty accounts for the correlation coefficient of 1 between FF_{tot} and FF_{geo} as described in Supplementary Information section 1), that is, still 65% greater than previous 3D inverse model estimates.

By mass balance, the microbial source (23% smaller than the literature[3–8]) must account for the difference between global CH_4 emissions, FF_{tot}, and biomass burning. Thus, our results suggest an important shift in the current understanding of the global CH_4 budget towards a higher FF_{tot} component compensated by lower microbial emissions, but the recent temporal increases in microbial emissions have been substantially larger. The $\delta^{13}C$ mass balance approach cannot distinguish between natural and anthropogenic microbial sources. However, the magnitude of the anthropogenic microbial sources (agriculture, including livestock and rice production, and waste, including landfill) has historically been related to population growth, and Schaefer *et al.*[21] recently found it plausible that agriculture and waste account for some of the temporal CH_4 emissions increase based on $\delta^{13}C$. All estimated pre-industrial and present-day global CH_4 budget terms are summarized in Table 1.

Table 1 | $\delta^{13}C$-based source attribution means for different periods.

	0–1700 AD*	1985–2002 AD	2003–2013 AD
Total fossil fuels†	51 ± 20	211 ± 33	195 ± 32
Fossil fuel industries	0	161 ± 24	145 ± 23
Geological sources		51 ± 20	
Microbial	154 ± 19	330 ± 28	355 ± 27
Biomass burning	25 ± 5	43 ± 9	

Values are given as mean ± one standard deviation in units of teragrams of methane per year. The biomass burning ranges are those prescribed in the $\delta^{13}C$ mass balance.
*See text and Supplementary Information section 6.
†TM5 simulations of the latitudinal gradient and comparison with observations indicate a present-day FF_{tot} range of 150–200 Tg CH_4 yr^{-1} (see text and Supplementary section 7).

We further evaluated our $\delta^{13}C$-based source attribution by simulating global atmospheric CH_4 mole fractions using emission maps scaled by the $\delta^{13}C$-based source attribution, and transported by the 3D global atmospheric chemistry and transport model TM5[27]. The simulated and observed global north–south gradients of CH_4 at remote background sites of NOAA's Global Greenhouse Gas Reference Network[20] add a spatial source attribution constraint because the broad spatial distribution of some CH_4 source categories is relatively well known globally. On the basis of nine simulated scenarios, we find that FF_{tot} in the range 150–200 Tg CH_4 yr^{-1} is consistent with the observed global north–south gradient (Supplementary Information section 7).

Inferred natural-gas industry efficiency improvements are illustrated using the FER time series in Fig. 3, which is calculated as FF_{ind} emissions (FF_{tot} minus FF_{geo}) minus oil and coal emissions (bottom-up estimates including uncertainties[28]) divided by global dry gas production of natural gas[16] and accounting for the CH_4 content of natural gas (Supplementary Information section 1). Mean FER decreases from 7.6% in 1985 to 2.2% in 2013. Assuming 'traditional' $\delta^{13}C_{source}$ values for all sources (blue dashed lines) would yield negative mean FER in some years, which is impossible, thereby emphasizing the importance of the updated $\delta^{13}C_{source}$. Note that the oil and coal emissions used to estimate FER assume temporally constant oil and coal emission factors while production increased, that is, no efficiency

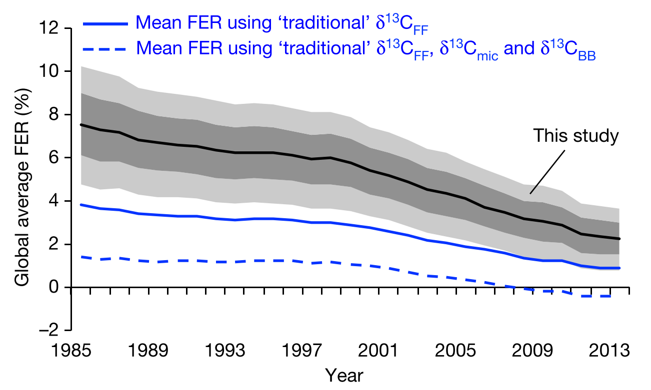


Figure 3 | Global FER long-term trend with mean values shown in solid black. Moving averages are shown; see Supplementary Fig. 18 for original mass balance results including inter-annual variability due to multiple components not accounted for in this model (see text). Dark and light grey bands mark the 25th/75th percentile and the 10th/90th percentile uncertainties, respectively. The dashed black line represents the linear trend of the means. Blue lines assume the mean $\delta^{13}C_{source}$ values from the literature (as in Fig. 1, blue values).

improvements in the oil and coal industries. Thus, the FER decline rate would be smaller in the case of noticeable oil and coal efficiency improvements. However, the impact of potential oil CH_4 emission reductions would be minor because oil contributes only on average 10% to FF_{ind} CH_4 emissions (Supplementary Fig. 10). Global coal CH_4 mitigation from reported projects, which may not be included in the bottom-up estimates amounts to only around 7% of global coal CH_4 emissions[1,29].

Our finding that FF_{tot} CH_4 emissions are 60%–110% higher than previous studies based on the most comprehensive global $\delta^{13}C_{source}$ database compiled so far represents a major adjustment in the global CH_4 budget, and this is consistent with the observed latitudinal CH_4 gradient. It agrees with a previous radiocarbon analysis[30] (167 ± 13 Tg CH_4 yr^{-1} FF_{tot}), which had so far been considered a "plausible re-estimate rather than a definitive revision"[30] of FF_{tot} owing to the model complexity involved. Our revised FF_{tot} emissions are compensated by lower microbial CH_4 emissions, and this is consistent with the palaeo-CH_4 budget (Supplementary Information section 6).

Accounting for previously neglected FF_{geo}, our correction of 20%–60% higher CH_4 emissions from natural gas, oil and coal production and use implies a greater potential for industry efficiency improvements to mitigate anthropogenic climate forcing. Yet, this study does not confirm an upward trend of FF_{ind} emissions in global CH_4 inventories[1,2] despite the large increase in natural gas, oil and coal production and use over the past three decades. Instead, this study finds that natural-gas CH_4 emissions per unit of production have declined from about 8% in the mid-1980 s to about 2% in the late 2000 s and early 2010 s. Natural-gas industry improvements associated with management practices, technology, and replacement of older equipment have been credited with reducing CH_4 leakage in the past[1]. The global observations used in our study confirm this trend, but the industry improvements have been offset by increased natural-gas production. Ongoing and future field studies at the level of natural-gas basins, facilities and components may help us to understand the contribution of individual leak types in order to reduce total natural-gas CH_4 emissions.

Received 25 April; accepted 22 August 2016.

1. US Environmental Protection Agency. *Summary Report: Global Anthropogenic Non-CO$_2$ Greenhouse Gas Emissions: 1990–2030* https://www3.epa.gov/climatechange/Downloads/EPAactivities/Summary_Global_NonCO2_Projections_Dec2012.pdf (Office of Atmospheric Programs, Climate Change Division, US EPA, 2012).
2. European Commission—Joint Research Centre (JRC)/Netherlands Environmental Assessment Agency (PBL). *Emission Database for Global Atmospheric Research (EDGAR)*, release version 4.2, http://edgar.jrc.ec.europa.eu (2014).
3. Bousquet, P. *et al.* Contribution of anthropogenic and natural sources to atmospheric methane variability. *Nature* **443,** 439–443 (2006).
4. Mikaloff Fletcher, S. E., Tans, P. P., Bruhwiler, L. M., Miller, J. B. & Heimann, M. CH_4 sources estimated from atmospheric observations of CH_4 and its $^{13}C/^{12}C$ isotopic ratios: 1. Inverse modeling of source processes. *Glob. Biogeochem. Cycles* **18,** GB4004 (2004).
5. Fraser, A. *et al.* Estimating regional methane surface fluxes: the relative importance of surface and GOSAT mole fraction measurements. *Atmos. Chem. Phys.* **13,** 5697–5713 (2013).
6. Wang, J. S. *et al.* A 3-D model analysis of the slowdown and interannual variability in the methane growth rate from 1988 to 1997. *Glob. Biogeochem. Cycles* **18,** GB3011 (2004).
7. Chen, Y.-H. & Prinn, R. G. Estimation of atmospheric methane emissions between 1996 and 2001 using a three-dimensional global chemical transport model. *J. Geophys. Res. Atmos.* **111,** D10307 (2006).
8. Bergamaschi, P. *et al.* Inverse modeling of global and regional CH_4 emissions using SCIAMACHY satellite retrievals. *J. Geophys. Res.* **114,** D22301 (2009).
9. Schwietzke, S., Griffin, W. M., Matthews, H. S. & Bruhwiler, L. M. P. Natural gas fugitive emissions rates constrained by global atmospheric methane and ethane. *Environ. Sci. Technol.* **48,** 7714–7722 (2014).
10. Quay, P. *et al.* The isotopic composition of atmospheric methane. *Glob. Biogeochem. Cycles* **13,** 445–461 (1999).
11. Kirschke, S. *et al.* Three decades of global methane sources and sinks. *Nat. Geosci.* **6,** 813–823 (2013).
12. Etiope, G., Lassey, K. R., Klusman, R. W. & Boschi, E. Reappraisal of the fossil methane budget and related emission from geologic sources. *Geophys. Res. Lett.* **35,** L09307 (2008).
13. Ciais, P. *et al.* in *Working Group I Contribution To The IPCC Fifth Assessment Report. Climate Change 2013—The Physical Science Basis* (eds Stocker, T. F. *et al.*) Ch. 6, 465–570 (Cambridge Univ. Press, 2013).
14. Sherwood, O., Schwietzke, S., Arling, V. & Etiope, G. *Global Inventory of Fossil and Non-fossil Methane $\delta^{13}C$ Source Signature Measurements for Improved Atmospheric Modeling* http://www.esrl.noaa.gov/gmd/ccgg/d13C-src-inv/ (NOAA/ESRL/GMD, 2016).
15. Alvarez, R. A., Pacala, S. W., Winebrake, J. J., Chameides, W. L. & Hamburg, S. P. Greater focus needed on methane leakage from natural gas infrastructure. *Proc. Natl Acad. Sci. USA* **109,** 6435–6440 (2012).
16. US Energy Information Administration *International Energy Statistics* database http://www.eia.gov/cfapps/ipdbproject/IEDIndex3.cfm ? tid=3&pid=3&aid=1 (2016).
17. Whiticar, M. J. in *Organic Geochemistry* (eds Durand, B. & Behar, F.) Vol. 16, 531–548 (Pergamon, 1990).
18. Milkov, A. V. Worldwide distribution and significance of secondary microbial methane formed during petroleum biodegradation in conventional reservoirs. *Org. Geochem.* **42,** 184–207 (2011).
19. Meslé, M., Dromart, G. & Oger, P. Microbial methanogenesis in subsurface oil and coal. *Res. Microbiol.* **164,** 959–972 (2013).
20. NOAA Earth System Research Laboratory Global Monitoring Division *Global Greenhouse Gas Reference Network* http://www.esrl.noaa.gov/gmd/ccgg/iadv (2016).
21. Schaefer, H. *et al.* A 21st century shift from fossil-fuel to biogenic methane emissions indicated by $^{13}CH_4$. *Science* **352,** 80–84 (2016).
22. US Environmental Protection Agency. *Inventory of U.S. Greenhouse Gas Emissions and Sinks: 1990–2014* https://www.epa.gov/ghgemissions/us-greenhouse-gas-inventory-report-1990-2014 (2016).
23. Randerson, J. T., Chen, Y., van der Werf, G. R., Rogers, B. M. & Morton, D. C. Global burned area and biomass burning emissions from small fires. *J. Geophys. Res.* **117,** G04012 (2012).
24. Sapart, C. J. *et al.* Natural and anthropogenic variations in methane sources during the past two millennia. *Nature* **490,** 85–88 (2012).
25. Mischler, J. A. *et al.* Carbon and hydrogen isotopic composition of methane over the last 1000 years. *Glob. Biogeochem. Cycles* **23,** GB4024 (2009).
26. Sowers, T. Atmospheric methane isotope records covering the Holocene period. *Quat. Sci. Rev.* **29,** 213–221 (2010).
27. Krol, M. *et al.* The two-way nested global chemistry-transport zoom model TM5: algorithm and applications. *Atmos. Chem. Phys.* **5,** 417–432 (2005).
28. Schwietzke, S., Griffin, W. M., Matthews, H. S. & Bruhwiler, L. M. P. Global bottom-up fossil fuel fugitive methane and ethane emissions inventory for atmospheric modeling. *ACS Sustain. Chem. Eng.* **2,** 1992–2001 (2014).
29. International Energy Agency *Coal Mine Methane in China: A Budding Asset with the Potential to Bloom* https://www.iea.org/publications/freepublications/publication/china_cmm_report.pdf (2009).
30. Lassey, K. R., Lowe, D. C. & Smith, A. M. The atmospheric cycling of radiomethane and the 'fossil fraction' of the methane source. *Atmos. Chem. Phys.* **7,** 2141–2149 (2007).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank J. Randerson, K. Johnson, A. Cole, C. Itle, T. Wirth, T. Capehart, S. Montzka, and R. Klusman for comments and discussions. We acknowledge M. Schoell and A. Ionescu for contributing $\delta^{13}C_{source}$ data. This research was supported by a National Research Council RAP fellowship and a CIRES IRP grant.

Author Contributions S.S. was responsible for study design, box-model development, analysis of TM5 results, manuscript preparation, and helped provide data. P.P.T. and J.B.M. helped with study design and improved the manuscript. O.A.S., G.E., E.J.D., S.E.M., V.A.A., B.H.V. and J.W.C.W. provided data and improved the manuscript. L.M.P.B. was responsible for TM5 modelling, helped with model analysis, and improved the manuscript.

Author Information The reconstructed atmospheric global average $\delta^{13}C_{atm}$ data for 1984–1998 are from ref. 21 and global $\delta^{13}C_{ice}$ data are from ref. 24. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.S. (stefan.schwietzke@noaa.gov).

Reviewer Information *Nature* thanks G. Allen, M. Heimann and the other anonymous reviewer(s) for their contribution to the peer review of this work.

LETTER

doi:10.1038/nature19365

Late Pleistocene climate drivers of early human migration

Axel Timmermann[1,2] & Tobias Friedrich[1]

On the basis of fossil and archaeological data it has been hypothesized that the exodus of *Homo sapiens* out of Africa and into Eurasia between ∼50–120 thousand years ago occurred in several orbitally paced migration episodes[1–4]. Crossing vegetated pluvial corridors from northeastern Africa into the Arabian Peninsula and the Levant and expanding further into Eurasia, Australia and the Americas, early *H. sapiens* experienced massive time-varying climate and sea level conditions on a variety of timescales. Hitherto it has remained difficult to quantify the effect of glacial- and millennial-scale climate variability on early human dispersal and evolution. Here we present results from a numerical human dispersal model, which is forced by spatiotemporal estimates of climate and sea level changes over the past 125 thousand years. The model simulates the overall dispersal of *H. sapiens* in close agreement with archaeological and fossil data and features prominent glacial migration waves across the Arabian Peninsula and the Levant region around 106–94, 89–73, 59–47 and 45–29 thousand years ago. The findings document that orbital-scale global climate swings played a key role in shaping Late Pleistocene global population distributions, whereas millennial-scale abrupt climate changes, associated with Dansgaard–Oeschger events, had a more limited regional effect.

Numerous studies[5–7] have postulated that human dispersal and evolution were a direct consequence of orbital-scale climate shifts (Fig. 1b–d) during the Late Pleistocene (126–11 thousand years ago (ka)) and the Holocene (11–0 ka). Every ∼21 thousand years decreased precession (Fig. 1a) and corresponding higher boreal summer insolation intensified rainfall in northern Africa, the Arabian Peninsula and the Levant[8], thus generating habitable savannah-type corridors[1,2] for *H. sapiens* and a possible exchange pathway between African and Eurasian populations, which in turn impacted the subsequent global dispersal pattern and gene flow of *H. sapiens* across Asia, Europe, Australia and into the Americas.

Elucidating the response of *H. sapiens* dispersal to past climate shifts has been hindered by the sparseness of palaeoenvironmental data in key regions[4] such as northeastern Africa, the Levant and the Arabian Peninsula, by regional uncertainties of global climate model simulations (see Methods, Extended Data Fig. 3), and by the prevailing dating uncertainties of fossil or archaeological records. Here we set out to quantify the effects of climate on human dispersal over the last glacial period, using a numerical reaction/diffusion human dispersal model (HDM, see Methods, Extended Data Fig. 1), which is forced by time-varying temperature, net primary production, desert fraction (Extended Data Figs 4–6) and sea level boundary conditions (Fig. 1f) obtained from a transient glacial/interglacial global earth system model simulation[9] covering the last 125 ka, an estimate of millennial-scale variability and sea level reconstructions[10], respectively (see Methods). The LOVECLIM climate model used here (see Methods)

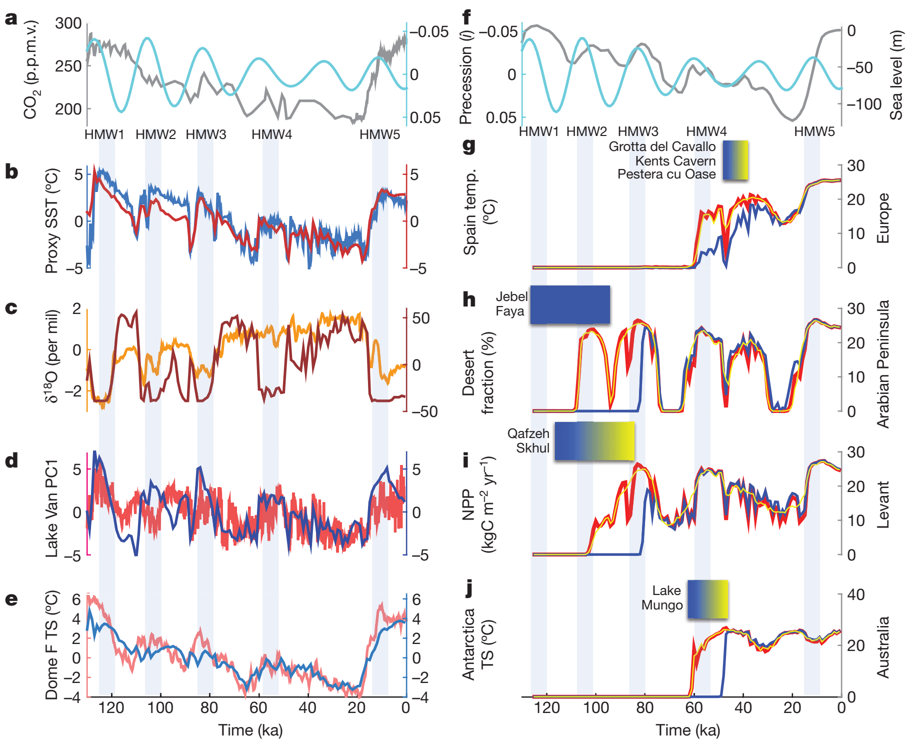


Figure 1 | Climate drivers. a, Precession (cyan) and CO_2 concentrations[24] (grey). p.p.m.v., parts per million by volume. **b**, Reconstructed[25] (blue) and simulated (red) North Atlantic/Spain temperature anomalies (°C) (see Methods). **c**, Israel (Soreq cave) speleothem[26] $\delta^{18}O$ (orange) and simulated Israel desert fraction anomalies (brown). **d**, Turkey hydroclimate reconstruction[27] (red) and simulated Turkey net primary production anomalies (NPP) (blue). **e**, Antarctic (Dome Fuji) reconstructed[28] (orange) and simulated (blue) surface temperature anomalies (TS). **f**, Precession and sea level[10]. **g–j**, Simulated human density (individuals per 100 km^2) for early (red), late (blue) and early-without-Dansgaard–Oeschger events (yellow) exit scenarios in Europe (9.5° E, 48.5° N), Arabian Peninsula (56.5° E, 21.5° N), Levant (37.5° E, 34.5° N) and Australia (144.5° E, 19.5° S). Blue (mixed blue/yellow) boxes indicate archaeological (and fossil) evidence for *H. sapiens*[18,20,29,30].

[1]International Pacific Research Center, University of Hawaii at Manoa, Honolulu, Hawaii 96822, USA. [2]Department of Oceanography, University of Hawaii at Manoa, Honolulu, Hawaii 96822, USA.

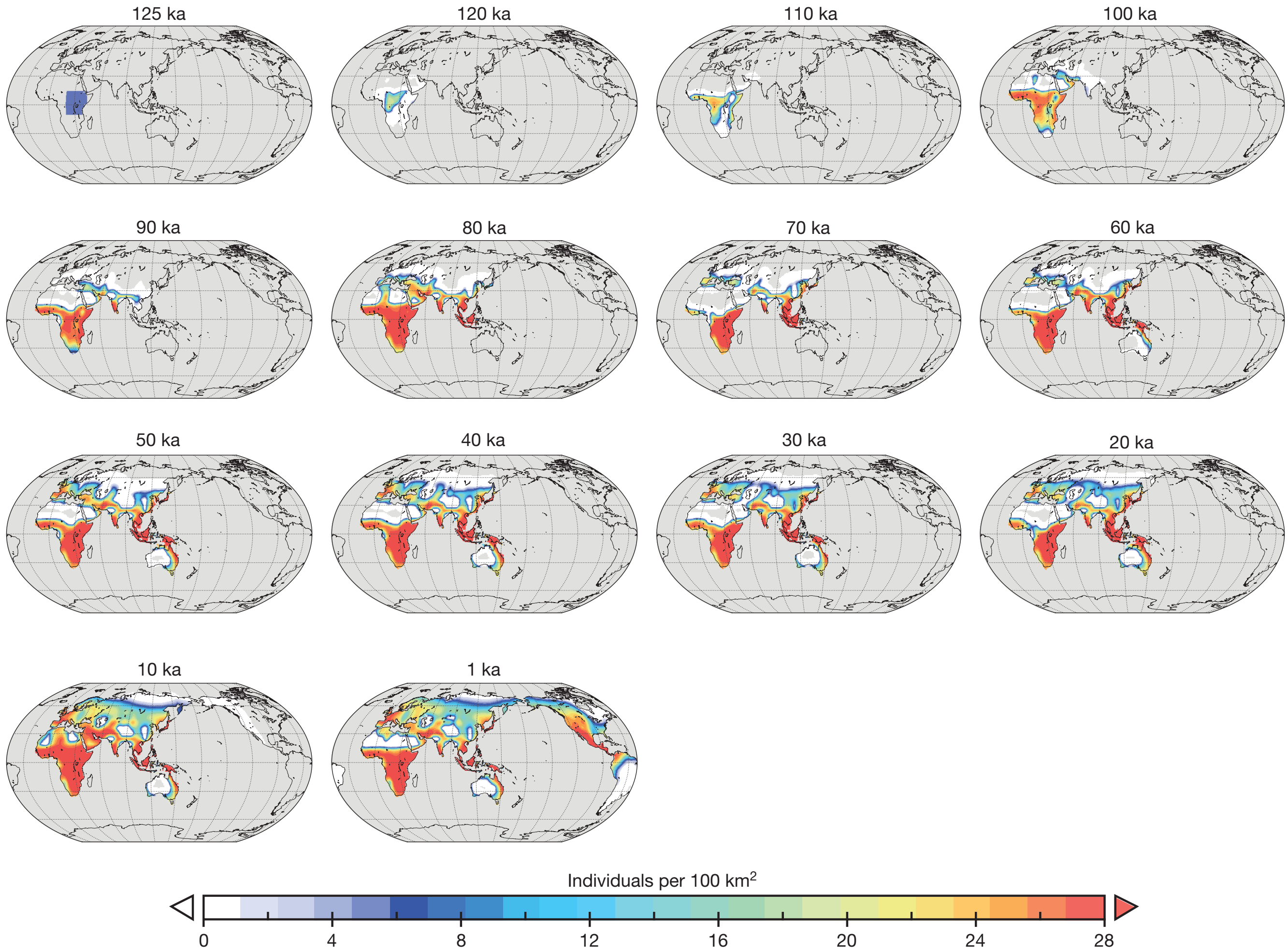


Figure 2 | Late Pleistocene human dispersal. Snapshots of the simulated evolution of human density (individuals per 100 km^2) over the past 125 thousand years using the parameters of scenario A (early exit) experiment (see Methods) with full climate (orbital- and millennial-scale) and sea level forcing and with human adaptation (see Methods).

simulates glacial temperature and hydroclimate variability in good agreement with some palaeoclimate proxy data[9] (Fig. 1b–e, Extended Data Fig. 2).

The first climate-forced HDM experiment (scenario A, Extended Data Tables 1, 2) covers the time evolution of the past 125 ka (Fig. 2, Supplementary Video 1) and accounts for a numerical representation of gradual human adaptation to environmental conditions (see Methods). Diffusing into vegetated regions, the first low-density migration wave of *H. sapiens* reaches the coastline of northeast Africa and the Bab-el-Mandeb around 115–106 ka (Figs. 1h, 2, Supplementary Video 1). Two rapid dispersals through the migration-favourable anomalously wet Arabian Peninsula- and Sinai-corridors (Fig. 1c, d, h, i, Extended Data Fig. 4) occur between 107–95 and 90–75 ka. Very low densities are simulated for Southern Europe from 95–72 ka (Fig. 2). The subsequent dry conditions during the period 71–60 ka (Marine Isotope Stage 4, MIS4) (Fig. 1c, d, Extended Data Figs 5, 6) cut off the exchange between the populations in northeastern Africa, and the rapidly eastward-spreading group in southern Asia (Fig. 2). This is in stark contrast to previous suggestions[11] of a very active migration corridor through the southern Arabian Peninsula during this time. During the subsequent precession minimum (increased boreal summer insolation) around 60–47 ka (Fig. 1a), simulated rainfall enhances net primary production in northern Africa, the Levant and the Arabian Peninsula (Extended Data Fig. 5) and a second prominent migration wave leaves northern Africa (Fig. 1h, i, 2). The onset of this prominent wave around 60 ka (Fig. 1g) coincides with the youngest estimates of the L3-haplogroup-based age range for the time to most recent common ancestor (TMRCA) of 79–60 ka[11,12] and the subsequent emergence of mitochondrial DNA (mtDNA) haplogroups M, N and R. Eventually, this wave merges with the Eurasian population and leads to a boost of population density across Europe (Fig. 1g) and southern Asia. Meanwhile, *H. sapiens* cross the sea-level-altered Indonesian archipelago and arrive in Papua New Guinea and Australia around 60 ka (Figs. 1j, 2). For the subsequent period from 60–30 ka (MIS3) the model simulates a continuous Africa/Eurasia exchange of *H. sapiens* through the Levant (Fig. 1i) and an additional wave (45–30 ka) across the Bab-el-Mandeb and through the Arabian Peninsula (Fig. 1h). The dispersal across the Levant region is further modulated by the millennial-scale drought/pluvial variability associated with Dansgaard–Oeschger stadial/interstadial transitions (Fig. 1d, i). Completing the simulated grand journey from Africa to the Americas, *H. sapiens* cross the Bering land bridge into North America during the short period from 14–10 ka. With rising sea levels the Bering land bridge gets inundated during the glacial termination (Supplementary Video 1), thus terminating the genographic connectivity between Asia and North America.

According to this early exit scenario the first *H. sapiens* arrived in Europe, India and Southeast Asia and southern China in low densities (<5 individuals per 100 km^2) already between 100–70 ka (see Supplementary Video 1, Fig. 3a). Whereas the simulated

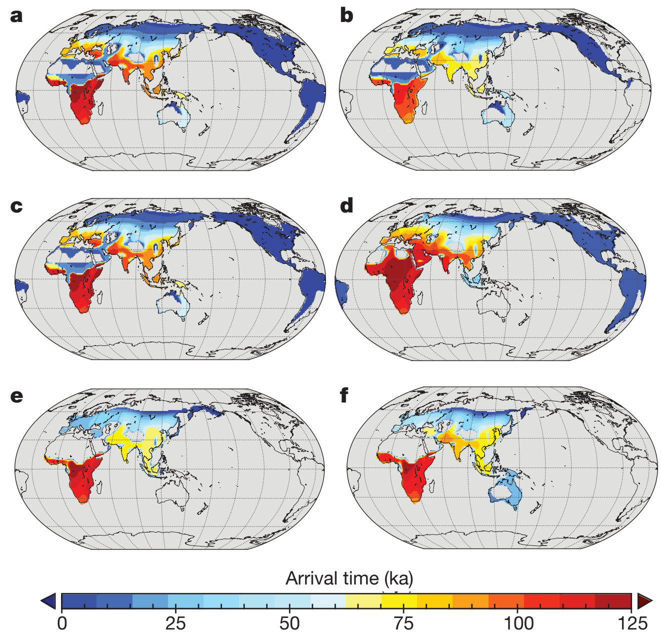


Figure 3 | Arrival Times for different dispersal scenarios. a–f, Time (ka) since last continuous human settlement for time-varying climate conditions following the transient scenario A simulation (early exit scenario) (**a**); scenario B (late exit scenario) (**b**); scenario A without Dansgaard–Oeschger variability (**c**); and for transient 125 ka simulations which use idealized constant climate forcing corresponding to 105 ka (**d**), 70 ka (**e**) and 21 ka (**f**) (see Extended Data Table 2 for more details).

early arrival of *H. sapiens* in Asia is consistent with previous findings[13–15], it appears to be at odds with a different interpretation of genetic and archaeological data[16]. Furthermore, the simulated low-density arrival of *H. sapiens* in Europe around 90–80 ka (Fig. 2) and the subsequent population increase from 60–50 ka challenges fossil and archaeological evidence[17] placing the European arrival of *H. sapiens* before or around 45 ka. A possible explanation for this large discrepancy between model and observational evidence could be that the small populations of *H. sapiens* arriving in southeastern Europe after Human Migration Window (HMW) 3 may have been assimilated by the prevalent Neanderthal population and that only the subsequent wave from the Levant (during HMW4, 60–47 ka) led to a gradual transition from a Neanderthal- to a *H. sapiens*-dominated population regime.

A second parameter scenario (scenario B, see Methods, Extended Data Tables 1, 2) was run with the HDM to quantify the effect of a potential late MIS5a/MIS4 exodus[11,16] on the subsequent population dynamics (Fig. 3b, Extended Data Fig. 7, Supplementary Video 2). According to this scenario, dispersal from central Africa to northeastern Africa is inhibited due the prevailing drought conditions in north Africa during 116–108.5 ka (MIS5d) and 91.5–84.5 ka (MIS5b) (Extended Data Figs 5–7, Supplementary Video 2) and the higher human temperature sensitivity chosen for this scenario. This period is followed by a very rapid dispersal from Africa into Eurasia across the Red Sea and Levant starting in 84 ka (Fig. 1h, i, Extended Data Fig. 7). For MIS4 low densities are simulated for the Levant, Arabian Peninsula, Southeast Asia and southeastern Europe (Extended Data Fig. 7). During HMW4 (60–47 ka, Fig. 1h) a second migration event takes place through the Arabian Peninsula and Levant and human densities increase in western Europe, the Middle East, India and Indonesia, partly due to local environmental conditions, partly due to an influx from other areas. This scenario resembles the late single southern exit model[11] and mtDNA evidence for haplogroup L3. Although this scenario explains qualitatively the reconstructed arrival times in India[16], Europe[18], Australia[19], and North America (Fig. 3b, Supplementary Video 2), it does not explain the early MIS5 presence of *H. sapiens* in the Levant[20], the Arabian Peninsula (Fig. 1g, h) and in southern China[15].

Both scenario A and B clearly reveal the impact of Dansgaard–Oeschger variability on the desert fraction and habitability of the Levant region (Extended Data Figs 5, 6, Fig. 1i). Whereas Dansgaard–Oeschger temperature and net primary production effects (Extended Data Figs 4, 5) on European, Mediterranean and north African population density are visible (Fig. 1g), their overall effect on global human dispersal and arrival times is negligible, as demonstrated by repeating scenario A without Dansgaard–Oeschger variability (see Methods) (Fig. 1g–i, yellow line, Fig. 3c).

To better understand the effect of glacial climate variability on the Late Pleistocene global human dispersal compared to non-climatic effects, we conducted three additional highly idealized sensitivity experiments (see Methods). For these 125 ka-long simulations we ignore the presence of glacial/interglacial climate variability and the orbital forcing and just prescribe perpetual climate conditions for 105 ka (MIS5c), 70 ka (MIS4), and 21 ka (Last Glacial Maximum, LGM). Comparing the times of last continuous human settlement simulated by scenario A simulation (Fig. 3a) with the corresponding maps for the perpetual MIS5c, MIS4 and LGM experiments (Fig. 3d, e, f), we find substantial differences in the peopling pattern during the Late Pleistocene, which can be understood either in terms of the differences in global climate conditions or in terms of sea level.

Here we presented a numerical modelling framework to quantify the effects of past spatiotemporal climate and sea level change on the global migration patterns of *H. sapiens*. Consistent with the recently proposed early onset multiple dispersal model[21] and supported by phenotype analyses of early human neurocranial geometries[22], we found evidence for multiple climate-mediated MIS3–5 dispersal and mixing waves across the Africa/Asian nexus. Precession forcing served as a key pacemaker for these glacial events which occurred around 106–94 ka (HMW2), 89–73 ka (HMW3) and 59–47 ka (HMW4) and 45–29 ka and which may have also contributed to potential gene flow back into Africa[21]. The large 59–47 ka dispersal event through the Arabian Peninsula, simulated by the HDM as well as by a recent demographic model[23], probably left the most prominent genetic traces in the genome of non-African *H. sapiens*, thus linking it to the TMRCA, the main gene flow patterns and estimated ages of mtDNA markers M and N and of the Y chromosome haplogroups M174, M130 and M89. Whether mtDNA evidence can be used to unequivocally distinguish between the orbital pulsation scenario and the late single-exodus scenario needs to be further explored. In addition to the orbital-scale pacing of human dispersal, we found modelling evidence for the impact of millennial-scale variability on regional population densities (Fig. 1g, h). However, according to our simulations the effects of abrupt climate change on global population dynamics and the first arrival times were negligible (Fig. 3c).

Consistent with a plethora of recent studies[14,15,21], our early exit climate/human dispersal simulation reproduces an early MIS5 exodus of *H. sapiens* out of Africa and a rapid dispersal along the southern rim of Asia into southern China and eventually into Australia during MIS3 and MIS4 (Supplementary Video 1). Our model simulation also shows an almost synchronous early arrival in southern China and in Europe around 90–80 ka. Given this plausible scenario, it is perplexing that the first *H. sapiens* fossil in southern China pre-dates the oldest discovered fossils of *H. sapiens* in Europe by about 35–40 thousand years. This discrepancy could be reconciled by assuming that the northern route into Europe was much more influenced by the biological and cultural interaction between *H. sapiens* and Neanderthals[17] than the southern route into Asia.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 7 January; accepted 12 August 2016.
Published online 21 September 2016.

1. Jennings, R. P. *et al.* The greening of Arabia: Multiple opportunities for human occupation of the Arabian Peninsula during the Late Pleistocene inferred from an ensemble of climate model simulations. *Quat. Int.* **382,** 181–199 (2015).
2. Parton, A. *et al.* Orbital-scale climate variability in Arabia as a potential motor for human dispersals. *Quat. Int.* **382,** 82–97 (2015).
3. Larrasoaña, J. C., Roberts, A. P. & Rohling, E. J. Dynamics of green Sahara periods and their role in hominin evolution. *PLoS One* **8,** e76514 (2013).
4. Breeze, P. S. *et al.* Palaeohydrological corridors for hominin dispersals in the Middle East similar to 250–70,000 years ago. *Quat. Sci. Rev.* **144,** 155–185 (2016).
5. Carto, S. L., Weaver, A. J., Hetherington, R., Lam, Y. & Wiebe, E. C. Out of Africa and into an ice age: on the role of global climate change in the late Pleistocene migration of early modern humans out of Africa. *J. Hum. Evol.* **56,** 139–151 (2009).
6. Castañeda, I. S. *et al.* Wet phases in the Sahara/Sahel region and human migration patterns in North Africa. *Proc. Natl Acad. Sci. USA* **106,** 20159–20163 (2009).
7. Scholz, C. A. *et al.* East African megadroughts between 135 and 75 thousand years ago and bearing on early-modern human origins. *Proc. Natl Acad. Sci. USA* **104,** 16416–16421 (2007).
8. Frumkin, A., Bar-Yosef, O. & Schwarcz, H. P. Possible paleohydrologic and paleoclimatic effects on hominin migration and occupation of the Levantine Middle Paleolithic. *J. Hum. Evol.* **60,** 437–451 (2011).
9. Timmermann, A. *et al.* Modeling Obliquity and CO_2 Effects on Southern Hemisphere Climate during the Past 408 ka. *J. Clim.* **27,** 1863–1875 (2014).
10. Waelbroeck, C. *et al.* Sea-level and deep water temperature changes derived from benthic foraminifera isotopic records. *Quat. Sci. Rev.* **21,** 295–305 (2002).
11. Oppenheimer, S. Out-of-Africa, the peopling of continents and islands: tracing uniparental gene trees across the map. Phil. *Trans. R. Soc. Lond. B* **367,** 770–784 (2012).
12. Soares, P. *et al.* The Expansion of mtDNA Haplogroup L3 within and out of Africa. *Mol. Biol. Evol.* **29,** 915–927 (2012).
13. Petraglia, M. *et al.* Middle Paleolithic assemblages from the Indian subcontinent before and after the Toba super-eruption. *Science* **317,** 114–116 (2007).
14. Petraglia, M. D., Haslam, M., Fuller, D. Q., Boivin, N. & Clarkson, C. Out of Africa: new hypotheses and evidence for the dispersal of *H. sapiens* along the Indian Ocean rim. *Ann. Hum. Biol.* **37,** 288–311 (2010).
15. Liu, W. *et al.* The earliest unequivocally modern humans in southern China. *Nature* **526,** 696–699 (2015).
16. Mellars, P., Gori, K. C., Carr, M., Soares, P. A. & Richards, M. B. Genetic and archaeological perspectives on the initial modern human colonization of southern Asia. *Proc. Natl Acad. Sci. USA* **110,** 10699–10704 (2013).
17. Benazzi, S. *et al.* Early dispersal of modern humans in Europe and implications for Neanderthal behaviour. *Nature* **479,** 525–528 (2011).
18. Fu, Q. *et al.* An early modern human from Romania with a recent Neanderthal ancestor. *Nature* **524,** 216–219 (2015).
19. Clarkson, C. *et al.* The archaeology, chronology and stratigraphy of Madjedbebe (Malakunanja II): A site in northern Australia with early occupation. *J. Hum. Evol.* **83,** 46–64 (2015).
20. Grün, R. *et al.* U-series and ESR analyses of bones and teeth relating to the human burials from Skhul. *J. Hum. Evol.* **49,** 316–334 (2005).
21. Groucutt, H. S. *et al.* Rethinking the dispersal of *H. sapiens* out of Africa. *Evol. Anthropol.* **24,** 149–164 (2015).
22. Gunz, P. *et al.* Early modern human diversity suggests subdivided population structure and a complex out-of-Africa scenario. *Proc. Natl Acad. Sci. USA* **106,** 6094–6098 (2009).
23. Eriksson, A. *et al.* Late Pleistocene climate change and the global expansion of anatomically modern humans. *Proc. Natl Acad. Sci. USA* **109,** 16089–16094 (2012).
24. Lüthi, D. *et al.* High-resolution carbon dioxide concentration record 650,000–800,000 years before present. *Nature* **453,** 379–382 (2008).
25. Martrat, B. *et al.* Four climate cycles of recurring deep and surface water destabilizations on the Iberian margin. *Science* **317,** 502–507 (2007).
26. Bar-Matthews, M., Ayalon, A., Kaufman, A. & Wasserburg, G. J. The Eastern Mediterranean paleoclimate as a reflection of regional events: Soreq cave, Israel. *Earth Planet. Sci. Lett.* **166,** 85–95 (1999).
27. Stockhecke, M. *et al.* Millennial to orbital-scale variations of drought intensity in the Eastern Mediterranean. *Quat. Sci. Rev.* **133,** 77–95 (2016).
28. Kawamura, K. *et al.* Northern Hemisphere forcing of climatic cycles in Antarctica over the past 360,000 years. *Nature* **448,** 912–916 (2007).
29. Armitage, S. J. *et al.* The southern route "out of Africa": evidence for an early expansion of modern humans into Arabia. *Science* **331,** 453–456 (2011).
30. Higham, T. *et al.* The earliest evidence for anatomically modern humans in northwestern Europe. *Nature* **479,** 521–524 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank S. Feakins, M. Segschneider and Y. Chikamoto for discussions and L. Menviel for providing the data of the LOVECLIM Dansgaard-Oeschger hindcast experiment, A. Ganopolski for providing the ice-sheet forcing from CLIMBER and M. Tigchelaar for providing the PMIP3 model data. A.T. is supported through the US NSF (grants 1341311, 1400914).

Author Contributions A.T. designed the research study, wrote the numerical model code for the human dispersal model, conducted the human dispersal numerical experiments and wrote the paper. T.F. ran the transient climate model simulation, conducted the model/proxy data comparison and contributed to the interpretation of the data.

Author Information The climate model and human dispersal model data are available on http://apdrc.soest.hawaii.edu/projects/HDM. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.T. (axel@hawaii.edu).

Reviewer Information *Nature* thanks P. deMenocal, R. Jennings, M. Petraglia and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

HDM equations and climate forcing. To simulate human migration out of Africa and its subsequent global dispersal under time-varying climate conditions we employ a reaction/diffusion model for human density $\rho(x,y,t)$ in each grid point (x,y) and at time t. Similar models, also known as modified Fisher–Kolmogorov or Fisher–Skellam models have been used in previous idealized human dispersal modelling studies[31,32]. In our HDM the reaction terms (growth G and mortality M) are dependent on three key vegetation, climate parameters: net primary productivity $N(x,y,t)$, which controls the availability of carbon-based food sources, desert fraction $d(x,y,t)$ and temperature $T(x,y,t)$, which are interpreted here as key climate stressors that determine the mortality of *H. sapiens*. Furthermore, a quadratic mortality term is included to avoid exponential growth during the simulation. The governing equation for $\rho(x,y,t)$ reads:

$$\frac{\partial \rho}{\partial t} = K\nabla^2\rho + \Gamma[N(x,y,t)]\rho - M[d(x,y,t), T(x,y,t)]\rho - \alpha\rho^2 \qquad (1)$$

The growth rate of human density is parameterized as a function of net primary productivity $N(x,y,t)$ in terms of $\Gamma(N) = \gamma\{0.5\tanh[(N - N_c)/N_w] + 0.4\}$. Assuming that the early human diet consisted largely of grazing mammals, this equation captures the notion that the primary productivity in each model grid box is linked to the edible biomass. A net primary productivity N_c is required in each grid box to maintain human population. We make the assumption that the mortality rate in each grid box is strongly controlled by the desert fraction d and the annual mean surface temperatures T. Climates with temperature below T_c will abruptly increase human mortality. Furthermore, we include a dependence of M on the desert fraction d. Human mortality will increase rapidly—a result of sparse food and water resources—if the desert fraction increases beyond d_c. Both factors are parameterized as: $M(d,t) = \mu \max\{1 + \tanh[(d - d_c)/d_w], 1.67[1 - \tanh[(T - T_c)/T_w]]\}$. The explicit equations used here for growth and mortality differ from the simplified logistic growth law used in previous studies[33].

Coastal express. For coastal points, we assume an alongshore advection with a mean alongshore velocity U_p (see Supplementary Table 1). The time evolution of ρ is then calculated from:

$$\frac{\partial \rho}{\partial t} = -U_p \cdot \nabla\rho + \Gamma[N(x,y,t)]\rho - M[d(x,y,t), T(x,y,t)]\rho - \alpha\rho^2$$

Accelerated migration along rivers[34] is not taken into account. The simulated human dispersal represents a strongly idealized scenario, which does not include competition or assimilation with *H. neanderthalensis*, *H. erectus* or Denisovans. To extend the model to a more realistic multi-actor framework would require more detailed information on the respective climate sensitivities and better observational constraints on the initial population densities.

Island hopping. Short-distance sea-faring is parameterized as a Gaussian decay away from the coast with a width of 1°.

Numerical implementation. The HDM is discretized using a first order upwind in time and second order central difference in space method. Our numerical implementation uses a $1° \times 1°$ horizontal grid with a time-step of 1 year. The physical model parameters chosen here are listed in Extended Data Table 1.

Orbital-scale simulation. To study the effects of slowly evolving glacial boundary conditions on the climate system and human dispersal we conducted a transient glacial–interglacial climate model simulation with the earth system model LOVECLIM (abbreviated as SIM). The simulation SIM is based on the earth system model LOVECLIM[35], version 1.1. The atmospheric component of the coupled model LOVECLIM is ECBilt[36], a spectral T21, three-level model, based on quasi-geostrophic equations extended by estimates of ageostrophic terms. The model contains a full hydrological cycle, which is closed over land by a bucket model for soil moisture and a runoff scheme. Diabatic heating due to radiative fluxes, the release of latent heat and the exchange of sensible heat with the surface are parameterized. The ocean-sea ice component of LOVECLIM, CLIO[37] consists of a free-surface Ocean General Circulation Model with $3° \times 3°$ resolution coupled to a thermodynamic-dynamic sea ice model. Coupling between atmosphere and ocean is done via the exchange of freshwater and heat fluxes, rather than by virtual salt fluxes. The terrestrial vegetation module of LOVECLIM, VECODE[38], computes the annual mean evolution of the vegetation cover based on annual mean values of several climatic variables.

Orbital-scale time-evolving ice-sheet boundary conditions in SIM are prescribed by changing ice-sheet orography and surface albedo. The corresponding anomalies were derived from the time-dependent ice-sheet reconstruction obtained from the CLIMBER, 2b earth system model of intermediate complexity[39,40]. Also, in LOVECLIM, the vegetation mask is adjusted to reflect time-evolving changes in ice-sheet-covered areas. Time-varying atmospheric greenhouse gas concentrations are prescribed in the model following greenhouse gas measurements from the EPICA DOME C ice core[24]. Another important forcing considered here is orbitally induced insolation variations that are calculated from the algorithm of ref. 41. We employ an acceleration technique, which compresses the time-varying external boundary conditions by a factor of 5. Instead of running the coupled model for the entire period of 408,000 years[42,43] the model experiment is 81,600 years long, while covering the entire forcing history of the last 408 thousand years. The acceleration technique is based on the assumption of relatively fast equilibration of surface variables to externally driven slow climate change. Through previous experimentation[42] we have found that an acceleration factor of 5 is appropriate for the tasks envisioned here. Whereas, the climate model run follows closely the methodology of ref. 9, the current simulation uses a higher climate sensitivity, which amounts to ~4 °C per CO_2 doubling. The result is a more realistic glacial/interglacial amplitude in surface temperatures compared to palaeo proxy data (see below).

To validate the model against palaeo proxy data we analyse 140 ka of the SIM model simulation (see below). This model simulation does not include the effects of millennial-scale variability associated with Dansgaard–Oeschger and Heinrich events. This will be added through a secondary model/data-based-procedure (see below).

While the full climate model simulation SIM covers 408 thousand years, only the last 125 thousand years are used to force the human dispersal model (see schematic Extended Data Fig. 1). The variables that are used as part of the climate forcing of the HDM are the simulated changes in temperature $[T_{orb}(x,y,t)]$, net primary productivity $[N_{orb}(x,y,t)]$ and desert fraction $[d_{orb}(x,y,t)]$.

Validation of orbital-scale LOVECLIM simulation. To compare the transient LOVECLIM simulation SIM with palaeo proxy data, we conduct an empirical orthogonal function (EOF) analysis of the simulated sea surface temperatures (SST) from 140–0 ka and compare the resulting leading EOF pattern and corresponding principal component with an EOF analysis conducted on 63 palaeo proxy SST proxy reconstructions[25,44–89]. The SST reconstructions[25,44–89] used for this analysis were required to cover the period from 140–10 ka. The location and the leading EOF pattern and corresponding principal component of these palaeo proxy data are shown in Extended Data Fig. 2. The model simulation reproduces both, the time evolution (Extended Data Fig. 2a) as well as the EOF pattern (Extended Data Fig. 2), in good agreement with the SST proxy data. We find higher EOF loadings for extratropical and some subtropical upwelling regions and somewhat damped EOF values for the tropical oceans. The resulting global mean SST time series which are based on the model and proxy EOF reconstructions (Extended Data Fig. 2c) exhibit a high degree of correlation and a similar magnitude for the transitions from the Last Glacial Maximum (LGM, 21 ka) to the early Holocene. However the magnitude of the Last Interglacial Ocean warming 130–120 ka is somewhat reduced in SIM, compared to the palaeo proxy reconstructions of global SST. There is also a reduction of the precessional (21 thousand year period) signal in the simulation relative to the palaeo data. Note that for this comparison we have made the assumption that annual mean SST in the model can be directly compared with a variety of SST proxies[25,44–89](such as alkenone, Mg/Ca and assemblage data). This requires that the proxies can be interpreted as an annual mean, which in some regions is not necessarily the case[90].

Other direct time series comparisons between simulated physical variables with palaeo proxy data are shown in Fig. 1. We find for instance that the magnitude as well as the timing of simulated surface temperatures over Antarctica in SIM matches the ice-core data well (Fig. 1e). Furthermore, the simulated hydroclimate variations in the Levant region bear close resemblance to orbital timescales with palaeo proxy reconstructions from speleothems (Fig. 1c) and lacustrine data (Fig. 1d).

To assess potential modelling uncertainties in hydroclimate for the selected time-slice of the LGM further we compare the simulated anomalies of the rainfall ratio between LGM and pre-industrial climate (relative to the pre-industrial values) in SIM with other CGCM LGM experiments conducted as part of the Palaeomodel Intercomparison Project, phase 3 (PMIP3). The results (Extended Data Fig. 3) clearly indicate that in spite of the fact that very similar boundary forcing conditions (greenhouse gas concentrations, ice-sheet topography and albedo and orbital forcing) are used, large modelling uncertainties exist in the simulated LGM rainfall patterns. Focusing on the particularly important human migration corridor of northern Africa, the LOVECLIM SIM experiment agrees well with the simulated drying in the CCSM4 and COSMOS-ASO simulations. In contrast the IPSL-CM5A-LR, GISS-E2-R and MPI-ESM-P models simulate significantly increased glacial rainfall over north Africa. For northern Europe, all CGCMs show

an overall drying for glacial conditions. The experimental design of PMIP3 does not include transient model experiments, so it is unclear at this stage, whether any of the other PMIP models would reproduce for example, the recorded hydroclimate variations over the past 125 thousand years in the Levant region (Fig. 1 d) in a qualitatively similar way as LOVECLIM.

Millennial-scale simulation and climate reconstruction. To account for the effects of millennial-scale Dansgaard–Oeschger (DO) and Heinrich events[91,92] on climate and the resulting changes in human dispersal, we reconstructed the corresponding anomalies of surface temperature $T'_m(x,y,t)$, net primary productivity $N'_m(x,y,t)$ and desert fraction $d'_m(x,y,t)$, (where the subscript m stands for the millennial-scale climate anomalies and the prime symbol denotes the deviation from the long-term mean) using a previously conducted transient LOVECLIM climate model simulation[93], (abbreviated as MIL). This simulation is a climate model hindcast of the period 50–30 ka, which includes both orbital scale forcings (ice-sheets, greenhouse gas concentrations and orbital changes) and freshwater-forced millennial-scale climate shifts associated with each observed DO and Heinrich event during this period. For this climate model simulation we calculated the regression patterns $P_i(x,y)$ between the simulated Iberian Margin SST $T_{ib}^{sim\prime}(t)$ and the corresponding spatiotemporal fields:

$$P_1(x,y)=\frac{\langle T_{ib}^{sim\prime}(t)T'_m(x,y,t)\rangle}{\langle (T_{ib}^{sim\prime})^2(t)\rangle}$$

$$P_2(x,y)=\frac{\langle T_{ib}^{sim\prime}(t)N'_m(x,y,t)\rangle}{\langle (T_{ib}^{sim\prime})^2(t)\rangle}$$

$$P_3(x,y)=\frac{\langle T_{ib}^{sim\prime}(t)d'_m(x,y,t)\rangle}{\langle (T_{ib}^{sim\prime})^2(t)\rangle}$$

To reconstruct the millennial-scale anomalies $[T^*_{m'}(x,y,t), N^*_{m'}(x,y,t), d^*_{m'}(x,y,t)]$ for the entire HDM period 130–0 ka, we multiply the model-based regression patterns with the observed high-pass-filtered Iberian margin SST variations[25] $T_{ib}^{obs\prime}$ that is, $T^*_{m'}(x,y,t)=T_{ib}^{obs\prime}(t)\ P_1(x,y)$; $N^*_{m'}(x,y,t)=T_{ib}^{obs\prime}(t)\ P_2(x,y)$; $d^*_{m'}(x,y,t)=T_{ib}^{obs\prime}(t)\ P_3(x,y)$. The resulting millennial-scale anomaly maps were then added back to the corresponding fields of the orbital scale SIM simulation $[T_{orb}(x,y,t), N_{orb}(x,y,t), d_{orb}(x,y,t)]$. The comparison between reconstructed and observed Iberian Margin SST variability shows an excellent agreement, on both the orbital and millennial timescale (Fig. 1b). Furthermore, we find a good agreement on both timescales between reconstructed and simulated hydroclimate variations in the Levant region (Fig. 1d).

Climate forcing for HDM simulations. The full climate forcing (temperature, net primary productivity and desert fraction) of the HDM is provided by adding the directly simulated LOVECLIM variables from SIM $T_{orb}(x,y,t)$, $N_{orb}(x,y,t)$, $d_{orb}(x,y,t)$ to the reconstructed millennial-scale anomalies $T^*_{m'}(x,y,t)$, $N^*_{m'}(x,y,t)$, $d^*_{m'}(x,y,t)$, obtained from model/data-based linear regression reconstruction, which uses model-derived patterns of millennial-scale variability $P_1(x,y)$, $P_2(x,y)$, $P_3(x,y)$ from MIL, and a time series $T_{ib}^{obs\prime}$ of observed millennial-scale variability from an Iberian Margin sediment proxy SST reconstruction. This yields: $T(x,y,t)=T_{orb}(x,y,t)+bT^*_{m'}(x,y,t)$, $N(x,y,t)=N_{orb}(x,y,t)+bN^*_{m'}(x,y,t)$, $d(x,y,t)=d_{orb}(x,y,t)+bd^*_{m'}(x,y,t)$, where b in the standard HDM simulation is set to 1 (see Extended Data Fig. 1). To further test the effect of DO variability on human dispersal, we also conducted one experiment with $b=0$ (Fig. 3c). We also include the time-varying coastline into the HDM climate driving fields $T(x,y,t)$, $N(x,y,t)$, $d(x,y,t)$.

The orbital and millennial-scale contributions to temperature, net primary productivity and desert fraction can again be retrieved statistically through an Empirical Orthgonal Function (EOF) analysis of the fields $T(x,y,t)$, $N(x,y,t)$, $d(x,y,t)$. The resulting EOF patterns and principal components are depicted in Extended Data Figs 4–6. Because of the linear operations conducted, the EOF patterns of millennial-scale variability for the respective variables are very similar to the regression patterns $P_1(x,y)$, $P_2(x,y)$, $P_3(x,y)$ (not shown).

Transient HDM simulations. To quantify the effects of climate variability and human dispersal processes on the dispersal of *H. sapiens* during the Late Pleistocene, the HDM is run in various parameter configurations and for a number of realistic and more idealized climate scenarios.

Initialization and scenarios. All experiments start from the same initial *H. sapiens* density distribution in central Africa (see Fig. 2, upper left for pattern), representing idealized initial conditions 125 ka. The results are essentially insensitive to moderate 15 degree latitudinal and longitudinal shifts of the initial *H. sapiens* across central Africa. These initial dates are chosen to study how the prominent MIS5 precessionally paced openings of savannah-type corridors in northeastern Africa during HMW2 and 3 (Fig. 1) would have promoted the subsequent human dispersal out of Africa.

By integrating Equation (1) forward in time and using different climate scenarios (see Extended Data Table 2) for Net primary productivity ($N(x,y,t)$), desert fraction ($d(x,y,t)$), surface temperature ($T(x,y,t)$) and land–sea distribution, we can determine the effects of orbital-scale and millennial-scale climate variability on the simulated arrival times of *H. sapiens* in various regions (Fig. 3) for the different scenarios. The following forcing scenarios are chosen (see Extended Data Tables 1, 2):

(1) The standard early exit control run (scenario A) (see parameters in Extended Data Table 1, left column) simulates the evolution of *H. sapiens* for the past 125 thousand years in response to time varying $T(x,y,t)$, $N(x,y,t)$, $d(x,y,t)$, which includes orbital-scale and millennial-scale variability. Furthermore, to capture an increased adaptability to climate stressors we reduce the values of T_c and increase those for d_c linearly during the simulation (see Extended Data Table 1). Furthermore, the mobility of *H. sapiens* increases as a function of time in terms of diffusion and coastal advection speeds (Extended Data Table 1, left column). The results are shown in Fig. 2 and Fig. 3a. The Late Holocene value of $K=42\,km^2\,year^{-1}$ in this simulation is in the range of previous estimates[31,33,94] for the Late Pleistocene *H. sapiens* and early- to mid-Holocene of $K=25–76\,km^2\,year^{-1}$. Our growth rate of 0.4% per year is a factor 3–4 smaller than previous estimates[32]. The simulated population range in scenario A for Europe during the LGM of 0.6 million individuals is about 3 times larger than a previous estimate[95].

(2) The late exit (scenario B) uses the same overall configuration as Scenario A (including same initial condition), but different parameter values (see Extended Data Table 1, right column). The results are shown in Extended Data Fig. 7 and Fig. 3b. Most importantly, this scenario uses a higher sensitivity of human mortality to temperatures and a smaller initial diffusion rate.

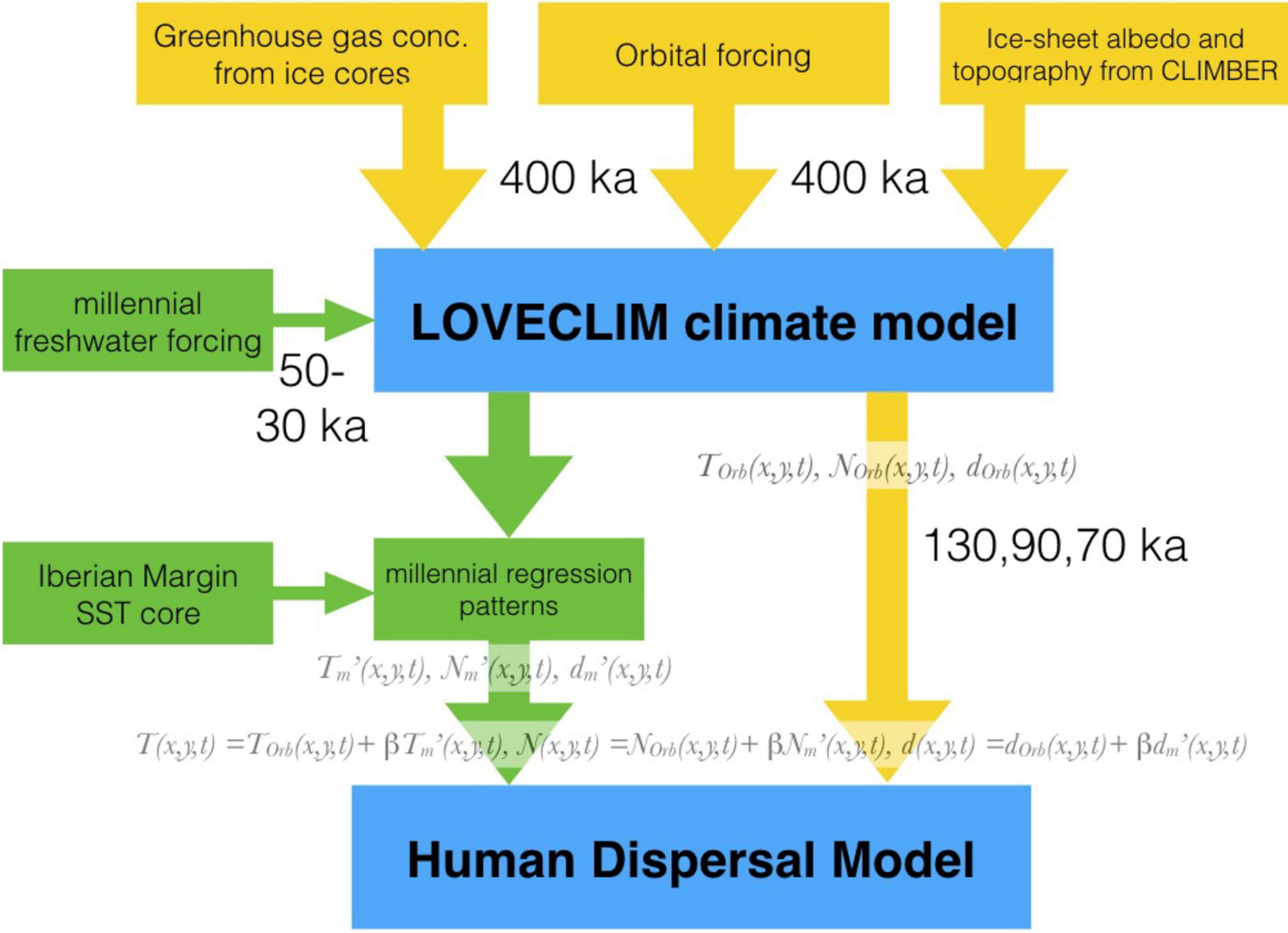
(3) A 125 ka simulation (no DO), based on scenario A, with only orbital-scale forcing $T_{orb}(x,y,t)$, $N_{orb}(x,y,t)$, $d_{orb}(x,y,t)$ from SIM and without millennial-scale variability associated with Dansgaard–Oeschger and Heinrich events. The arrival time results are shown in Fig. 3c and some representative time series in Fig. 1, yellow lines.

(4) Four 125 thousand-year-long HDM simulations using constant climate forcing $T_{orb}(x,y,t=t_c)$, $N_{orb}(x,y,t=t_c)$, $d_{orb}(x,y,t=t_c)$ with $t_c=105, 70, 21$ ka using the same parameters as in scenario A. The arrival time results are shown in Fig. 3d–f.

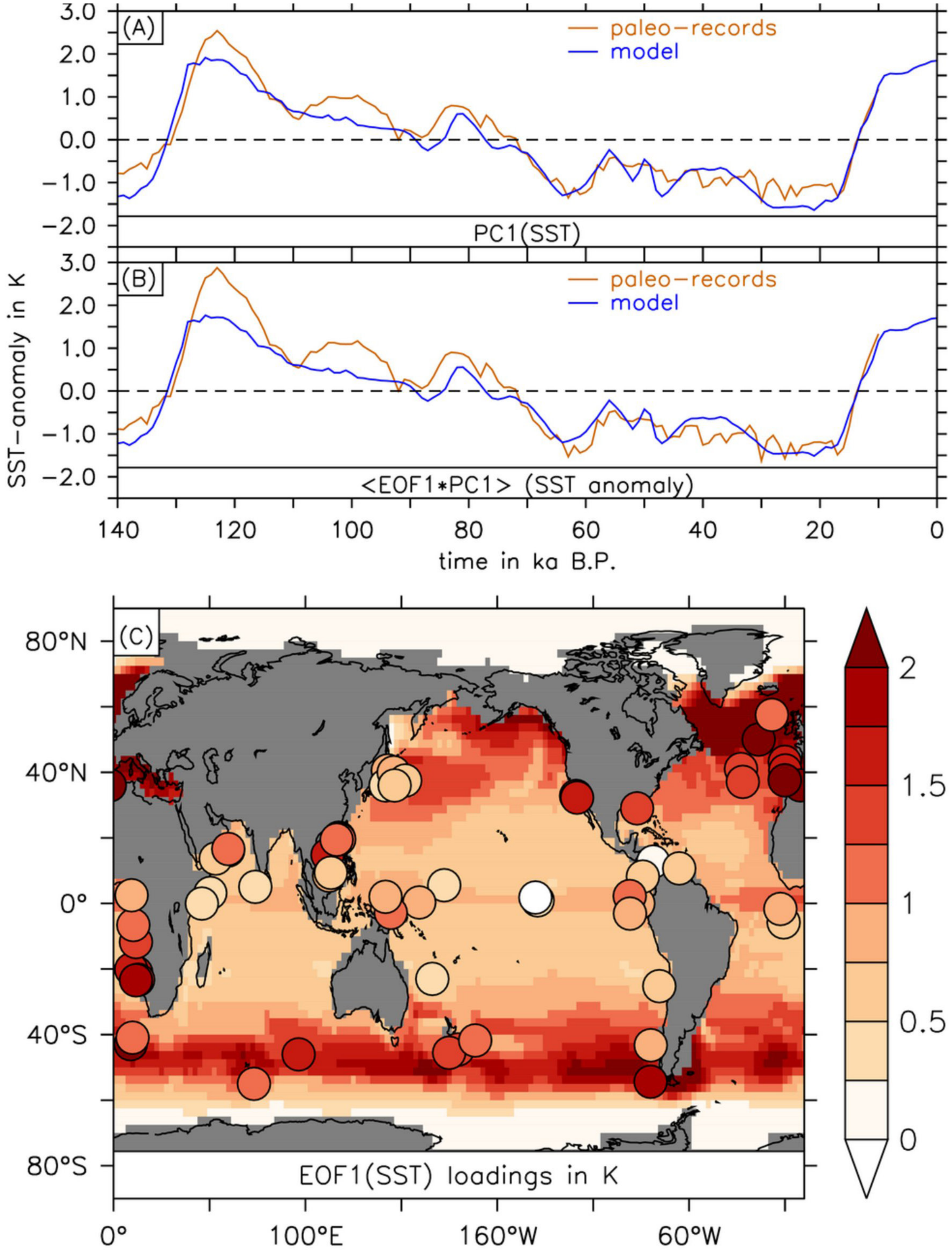
Code availability. The Matlab code of the human dispersion model is available upon request from the lead author.

31. Young, D. A. & Bettinger, R. L. Simulating the Global Human Expansion in the Late Pleistocene. *J. Archaeol. Sci.* **22,** 89–92 (1995).
32. Steele, J. Human dispersals: mathematical models and the archaeological record. *Hum. Biol.* **81,** 121–140 (2009).
33. Fort, J., Pujol, T. & Cavalli-Sforza, L. L. Palaeolithic populations and waves of advance (Human range expansions). *Camb. Archaeol. J.* **14,** 53–61 (2004).
34. Scerri, E. M. L., Groucutt, H. S., Jennings, R. P. & Petraglia, M. D. Unexpected technological heterogeneity in northern Arabia indicates complex Late Pleistocene demography at the gateway to Asia. *J. Hum. Evol.* **75,** 125–142 (2014).
35. Goosse, H. *et al.* Description of the Earth system model of intermediate complexity LOVECLIM version 1.2. *Geosci. Model Dev.* **3,** 603–633 (2010).
36. Opsteegh, J. D., Haarsma, R. J., Selten, F. M. & Kattenberg, A. ECBILT: a dynamic alternative to mixed boundary conditions in ocean models. *Tellus, Ser. A, Dyn. Meterol. Oceanogr.* **50,** 348–367 (1998).
37. Goosse, H. & Fichefet, T. Importance of ice-ocean interactions for the global ocean circulation: A model study. *Journal of Geophysical Research* **104,** 23337–23355 (1999).
38. Brovkin, V., Ganopolski, A. & Svirezhev, Y. A continuous climate-vegetation classification for use in climate-biosphere studies. *Ecol. Modell.* **101,** 251–261 (1997).
39. Ganopolski, A., Calov, R. & Claussen, M. Simulation of the last glacial cycle with a coupled climate ice-sheet model of intermediate complexity. *Clim. Past* **6,** 229–244 (2010).
40. Ganopolski, A. & Calov, R. The role of orbital forcing, carbon dioxide and regolith in 100 kyr glacial cycles. *Clim. Past* **7,** 2391–2411 (2011).
41. Berger, A. Long-term variations of caloric insolation resulting from the earth's orbital elements. *Quat. Res.* **9,** 139–167 (1978).
42. Timm, O. & Timmermann, A. Simulation of the last 21000 years using accelerated transient boundary conditions. *J. Clim.* **20,** 4377–4401 (2007).
43. Timm, O., Timmermann, A., Abe-Ouchi, A., Saito, F. & Segawa, T. On the definition of seasons in paleoclimate simulations with orbital forcing. *Paleoceanography* **23,** PA2221 (2008).
44. Lawrence, K., Herbert, T., Brown, C., Raymo, M. & Haywood, A. High-amplitude variations in North Atlantic sea surface temperature during the early Pliocene warm period. *Paleoceanography* **24,** (2009).
45. Naafs, B., Hefter, J. & Stein, R. Millennial-scale ice rafting events and Hudson Strait Heinrich(-like) Events during the late Pliocene and Pleistocene: a review. *Quat. Sci. Rev.* **80,** 1–28 (2013).
46. Etourneau, J., Martinez, P., Blanz, T. & Schneider, R. Pliocene-Pleistocene variability of upwelling activity, productivity, and nutrient cycling in the Benguela region. *Geology* **37,** 871–874 (2009).

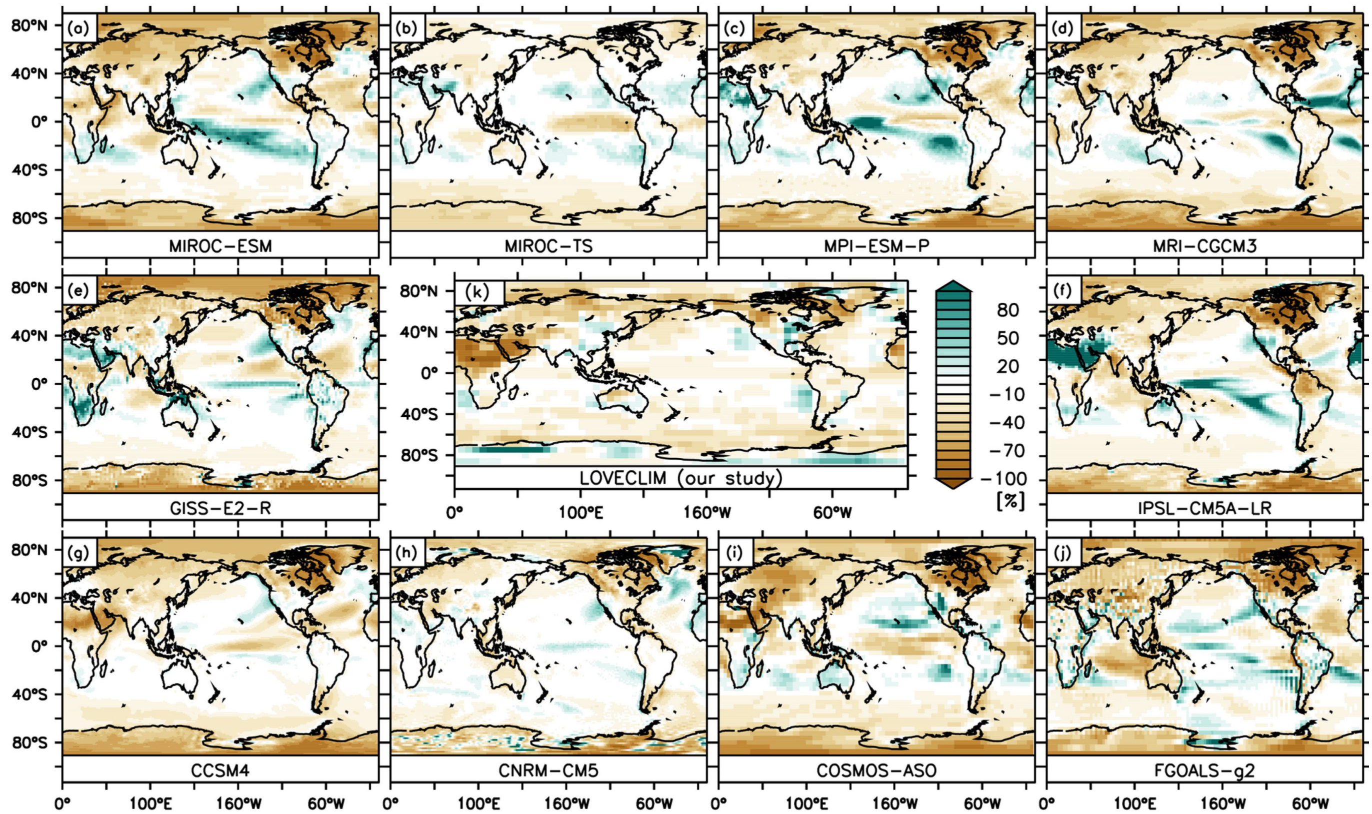
47. Herbert, T. D., Peterson, L. C., Lawrence, K. T. & Liu, Z. Tropical ocean temperatures over the past 3.5 million years. *Science* **328,** 1530–1534 (2010).
48. Li, L. *et al.* A 4-Ma record of thermal evolution in the tropical western Pacific and its implications on climate change. *Earth Planet. Sci. Lett.* **309,** 10–20 (2011).
49. de Garidel-Thoron, T., Rosenthal, Y., Bassinot, F. & Beaufort, L. Stable sea surface temperatures in the western Pacific warm pool over the past 1.75 million years. *Nature* **433,** 294–298 (2005).
50. Medina-Elizalde, M. & Lea, D. W. The mid-Pleistocene transition in the tropical Pacific. *Science* **310,** 1009–1012 (2005).
51. Russon, T. *et al.* Inter-hemispheric asymmetry in the early Pleistocene Pacific warm pool. *Geophys. Res. Lett.* **37,** (2010).
52. Pisias, N. & Rea, D. Late pleistocene paleoclimatology of the central equatorial pacific: sea surface response to the southeast trade winds. *Paleoceanography* **3,** 21–37 (1988).
53. Liu, Z. & Herbert, T. D. High-latitude influence on the eastern equatorial Pacific climate in the early Pleistocene epoch. *Nature* **427,** 720–723 (2004).
54. Schaefer, G. *et al.* Planktic foraminiferal and sea surface temperature record during the last 1 Myr across the Subtropical Front, Southwest Pacific. *Mar. Micropaleontol.* **54,** 191–212 (2005).
55. Hayward, B. *et al.* The effect of submerged plateaux on Pleistocene gyral circulation and sea-surface temperatures in the Southwest Pacific. *Global Planet. Change* **63,** 309–316 (2008).
56. Martínez-Garcia, A., Rosell-Melé, A., McClymont, E. L., Gersonde, R. & Haug, G. H. Subpolar link to the emergence of the modern equatorial Pacific cold tongue. *Science* **328,** 1550–1553 (2010).
57. Salgueiro, E. *et al.* Temperature and productivity changes off the western Iberian margin during the last 150 ky. *Quat. Sci. Rev.* **29,** 680–695 (2010).
58. Martinez, J., Mora, G. & Barrows, T. Paleoceanographic conditions in the western Caribbean Sea for the last 560 kyr as inferred from planktonic foraminifera. *Mar. Micropaleontol.* **64,** 177–188 (2007).
59. Voelker, A. H. & de Abreu, L. in *Abrupt Climate Change: Mechanisms, Patterns, and Impacts* Vol. 193 (eds Rashid, H., Polyak, L. & Mosley-Thompson, E.) (AGU, Geophysical Monograph Series, 2011).
60. Ziegler, M., Nurnberg, D., Karas, C., Tiedemann, R. & Lourens, L. Persistent summer expansion of the Atlantic Warm Pool during glacial abrupt cold events. *Nat. Geosci.* **1,** 601–605 (2008).
61. Herbert, T. D. & Schuffert, J. D. Vol. 165 1–9 (Proceedings of the Ocean Drilling Program, Scientific Results, College Station, TX (Ocean Drilling Program), 2000).
62. Calvo, E., Villanueva, J., Grimalt, J., Boelaert, A. & Labeyrie, L. New insights into the glacial latitudinal temperature gradients in the North Atlantic. Results from U-37(K ') sea surface temperatures and terrigenous inputs. *Earth Planet. Sci. Lett.* **188,** 509–519 (2001).
63. Weldeab, S., Lea, D. W., Schneider, R. R. & Andersen, N. 155,000 years of West African monsoon and ocean thermal evolution. *Science* **316,** 1303–1307 (2007).
64. Mix, A. C. & Fairbanks, R. G. North-Atlantic surface-ocean control of pleistocene deep-ocean circulation. *Earth Planet. Sci. Lett.* **73,** 231–243 (1985).
65. Martrat, B. *et al.* Abrupt temperature changes in the Western Mediterranean over the past 250,000 years. *Science* **306,** 1762–1765 (2004).
66. Schneider, R. R., Mueller, P. J. & Ruhland, G. Late quaternary surface circulation in the east Equatorial South-Atlantic - evidence from alkenone sea-surface temperatures. *Paleoceanography* **10,** 197–219 (1995).
67. Kirst, G., Schneider, R., Muller, P., von Storch, I. & Wefer, G. Late Quaternary temperature variability in the Benguela Current System derived from alkenones. *Quat. Res.* **52,** 92–103 (1999).
68. Nurnberg, D., Muller, A. & Schneider, R. Paleo-sea surface temperature calculations in the equatorial east Atlantic from Mg/Ca ratios in planktic foraminifera: A comparison to sea surface temperature estimates from U-37(K '), oxygen isotopes, and foraminiferal transfer function. *Paleoceanography* **15,** 124–134 (2000).
69. Budziak, D. in *Berichte aus dem Fachbereich Geowissenschaften der Universität Bremen* Vol. **170** 114pp (2000).
70. Bard, E., Rostek, F. & Sonzogni, C. Interhemispheric synchrony of the last deglaciation inferred from alkenone palaeothermometry. *Nature* **385,** 707–710 (1997).
71. Pelejero, C., Grimalt, J., Heilig, S., Kienast, M. & Wang, L. High-resolution U-37(K) temperature reconstructions in the South China Sea over the past 220 kyr. *Paleoceanography* **14,** 224–231 (1999).
72. Wei, G., Deng, W., Liu, Y. & Li, X. High-resolution sea surface temperature records derived from foraminiferal Mg/Ca ratios during the last 260 ka in the northern South China Sea. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **250,** 126–138 (2007).
73. Oppo, D. & Sun, Y. Amplitude and timing of sea-surface temperature change in the northern South China Sea: dynamic link to the East Asian monsoon. *Geology* **33,** 785–788 (2005).
74. Koizumi, I. & Yamamoto, H. Paleoceanographic evolution of North Pacific surface water off Japan during the past 150,000 years. *Mar. Micropaleontol.* **74,** 108–118 (2010).
75. Dyez, K. & Ravelo, A. Late Pleistocene tropical Pacific temperature sensitivity to radiative greenhouse gas forcing. *Geology* **41,** 23–26 (2013).
76. Tachikawa, K., Timmermann, A., Vidal, L., Sonzogni, C. & Timm, O. CO_2 radiative forcing and Intertropical Convergence Zone influences on western Pacific warm pool climate over the past 400 ka. *Quat. Sci. Rev.* **86,** 24–34 (2014).
77. Jasper, J. P., Hayes, J. M., Mix, A. C. & Prahl, F. G. Photosynthetic fractionation of 13C and concentrations of dissolved CO_2 in the central equatorial Pacific during the last 255,000 years. *Paleoceanography* **9,** 781–798 (1994).
78. Yu, P. *et al.* Influences of extratropical water masses on equatorial Pacific cold tongue variability during the past 160 ka as revealed by faunal evidence of planktic foraminifers. *J. Quaternary Sci.* **27,** 921–931 (2012).
79. Ho, S. *et al.* Sea surface temperature variability in the Pacific sector of the Southern Ocean over the past 700 kyr. *Paleoceanography* **27,** (2012).
80. Rincon-Martinez, D. *et al.* More humid interglacials in Ecuador during the past 500 kyr linked to latitudinal shifts of the equatorial front and the Intertropical Convergence Zone in the eastern tropical Pacific. *Paleoceanography* **25,** (2010).
81. Herbert, T. D. *et al.* Collapse of the California Current during glacial maxima linked to climate change on land. *Science* **293,** 71–76 (2001).
82. Lea, D. W., Pak, D. K. & Spero, H. J. Climate impact of late quaternary equatorial pacific sea surface temperature variations. *Science* **289,** 1719–1724 (2000).
83. Yamamato, M., Yamamuro, M. & Tanaka, Y. The California current system during the last 136,000 years: response of the North Pacific High to precessional forcing. *Quat. Sci. Rev.* **26,** 405–414 (2007).
84. Cortese, G., Abelmann, A. & Gersonde, R. A glacial warm water anomaly in the subantarctic Atlantic Ocean, near the Agulhas Retroflection. *Earth Planet. Sci. Lett.* **222,** 767–778 (2004).
85. Becquey, S. & Gersonde, R. A. 0.55-Ma paleotemperature record from the Subantarctic zone: Implications for Antarctic Circumpolar Current development. *Paleoceanography* **18,** (2003).
86. Sowers, T. *et al.* A 135,000-year Vostok-specmap common temporal framework. *Paleoceanography* **8,** 737–766 (1993).
87. Pichon, J. *et al.* Surface water temperature changes in the high latitudes of the southern hemisphere over the last glacial-interglacial cycle. *Paleoceanography* **7,** 289–318 (1992).
88. Sikes, E., Howard, W., Neil, H. & Volkman, J. Glacial-interglacial sea surface temperature changes across the subtropical front east of New Zealand based on alkenone unsaturation ratios and foraminiferal assemblages. *Paleoceanography* **17,** (2002).
89. Pahnke, K. & Sachs, J. Sea surface temperatures of southern midlatitudes 0–160 kyr BP. *Paleoceanography* **21,** (2006).
90. Timmermann, A., Sachs, J. & Timm, O. E. Assessing divergent SST behavior during the last 21 ka derived from alkenones and G. ruber-Mg/Ca in the equatorial Pacific. *Paleoceanography* **29,** 680–696 (2014).
91. Masson-Delmotte, V. *et al.* 383–464 (Cambridge University Press, 2013).
92. Sepulchre, P. *et al.* H4 abrupt event and late Neanderthal presence in Iberia. *Earth Planet. Sci. Lett.* **258,** 283–292 (2007).
93. Menviel, L., Timmermann, A., Friedrich, T. & England, M. H. Hindcasting the continuum of Dansgaard-Oeschger variability: mechanisms, patterns and timing. *Clim. Past* **10,** 63–77 (2014).
94. Pinhasi, R., Fort, J. & Ammerman, A. J. Tracing the origin and spread of agriculture in Europe. *PLoS Biol.* **3,** e410 (2005).
95. Tallavaara, M., Luoto, M., Korhonen, N., Järvinen, H. & Seppä, H. Human population dynamics in Europe over the Last Glacial Maximum. *Proc. Natl Acad. Sci. USA* **112,** 8232–8237 (2015).



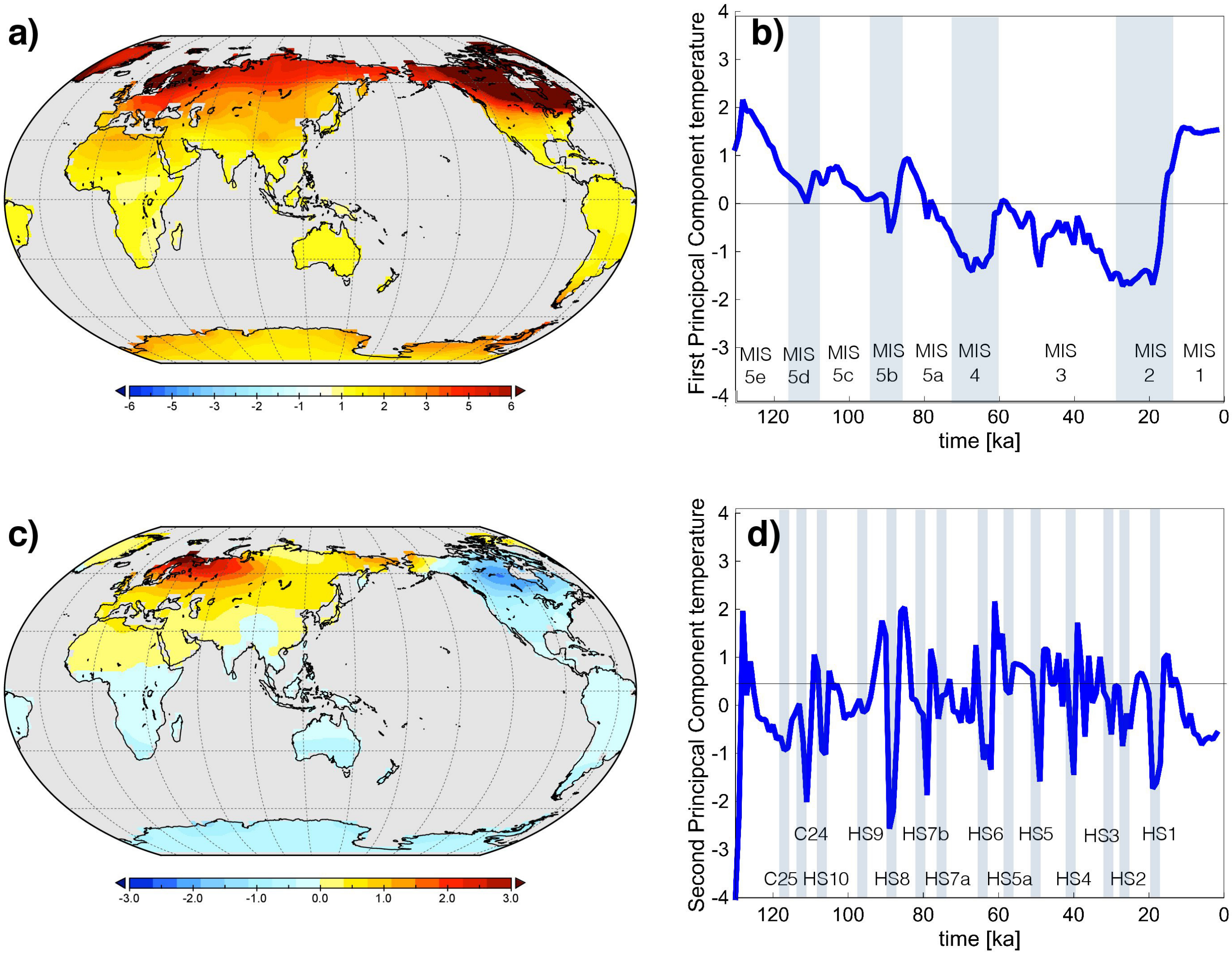
Extended Data Figure 1 | Schematics of modelling framework adopted for this study.



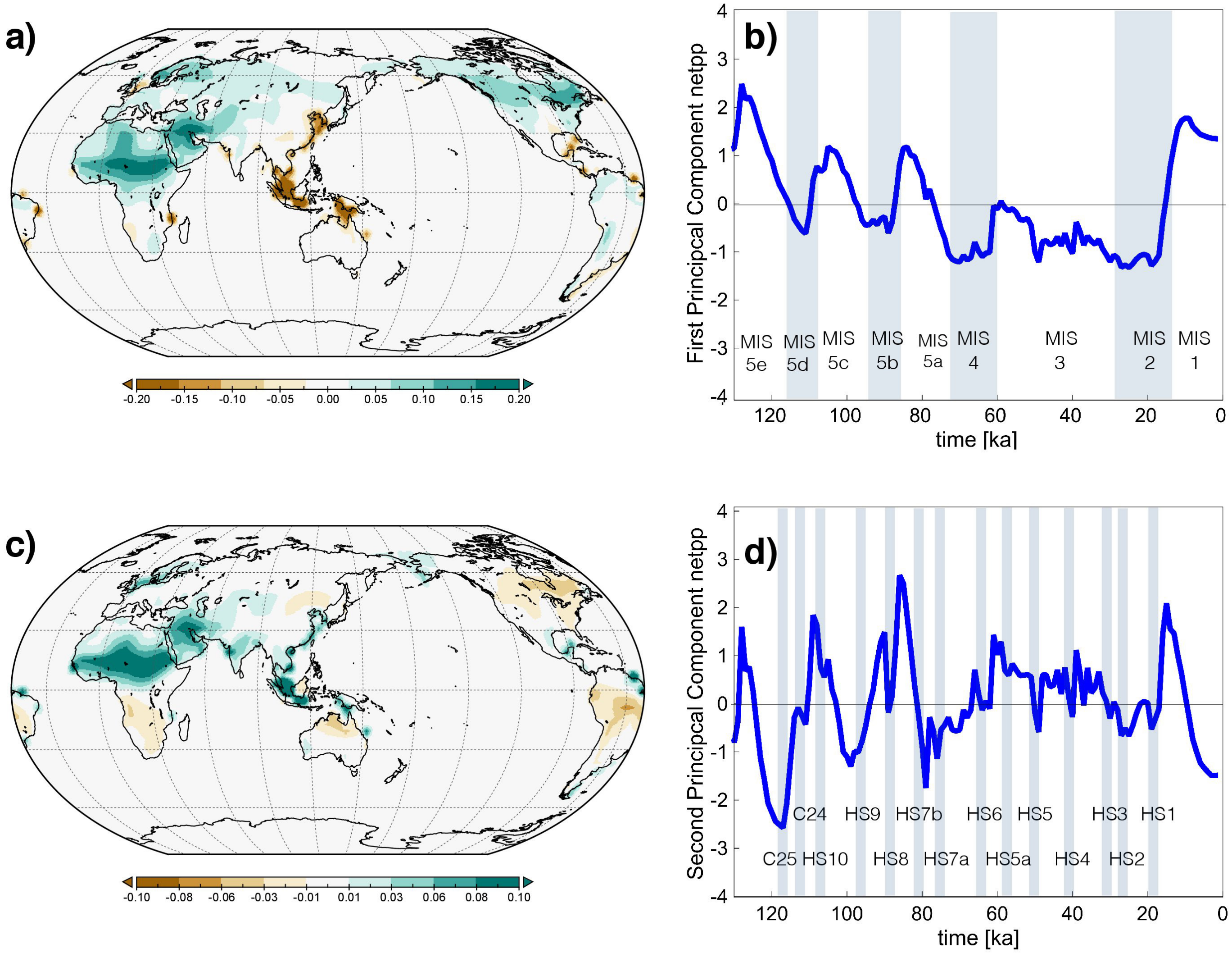
Extended Data Figure 2 | Validation of climate model simulation for temperature with palaeo sea surface temperature (SST) reconstructions. Pattern and temporal evaluation of leading Empirical Orthogonal Function (EOF1) of reconstructed and simulated SST. **a**, Principal components of the EOF1 (PC1) for SST from 63 palaeo-records[25,44–89] (orange) covering the period 140–10 ka and simulated SST (blue) using every model grid point. **b**, Globally-averaged SST anomaly (K) from EOF1-based reconstruction. Colours as in **a**. **c**, EOF1-pattern (K) for 63 palaeo records[25,44–89] (circles) and for simulated SST in global domain (shading).



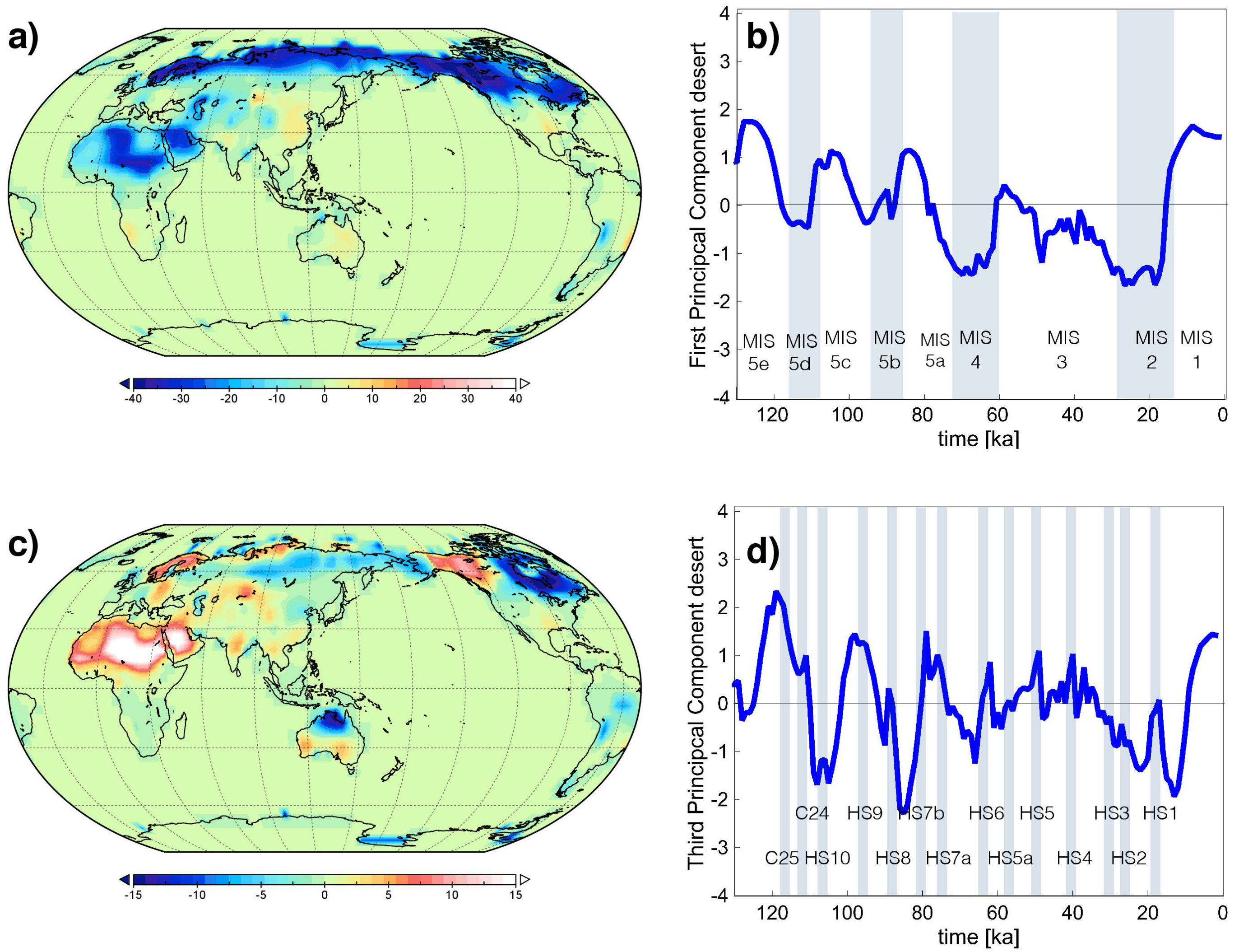
Extended Data Figure 3 | Comparison of LOVECLIM simulation with other PMIP3 CGCM Last Glacial Maximum simulations. a–**j**, Simulated annual mean rainfall differences (LGM versus pre-industrial) relative to the pre-industrial long-term annual mean rainfall (%) for ten different climate model simulations (MIROC-ESM (**a**), MIROC-TS (**b**), MPI-ESM-P (**c**), MRI-CGCM3 (**d**), GISS-E2-R (**e**), IPSL-CM5A-LR (**f**), CCSM4 (**g**), CNRM-CM5 (**h**), COSMOS-ASO (**i**) and FGOALS-g2 (**j**)) conducted as part of the Paleo Model Intercomparison Project, Phase 5 (PMIP5) (see Methods) and the LOVECLIM model (**k**) used here.



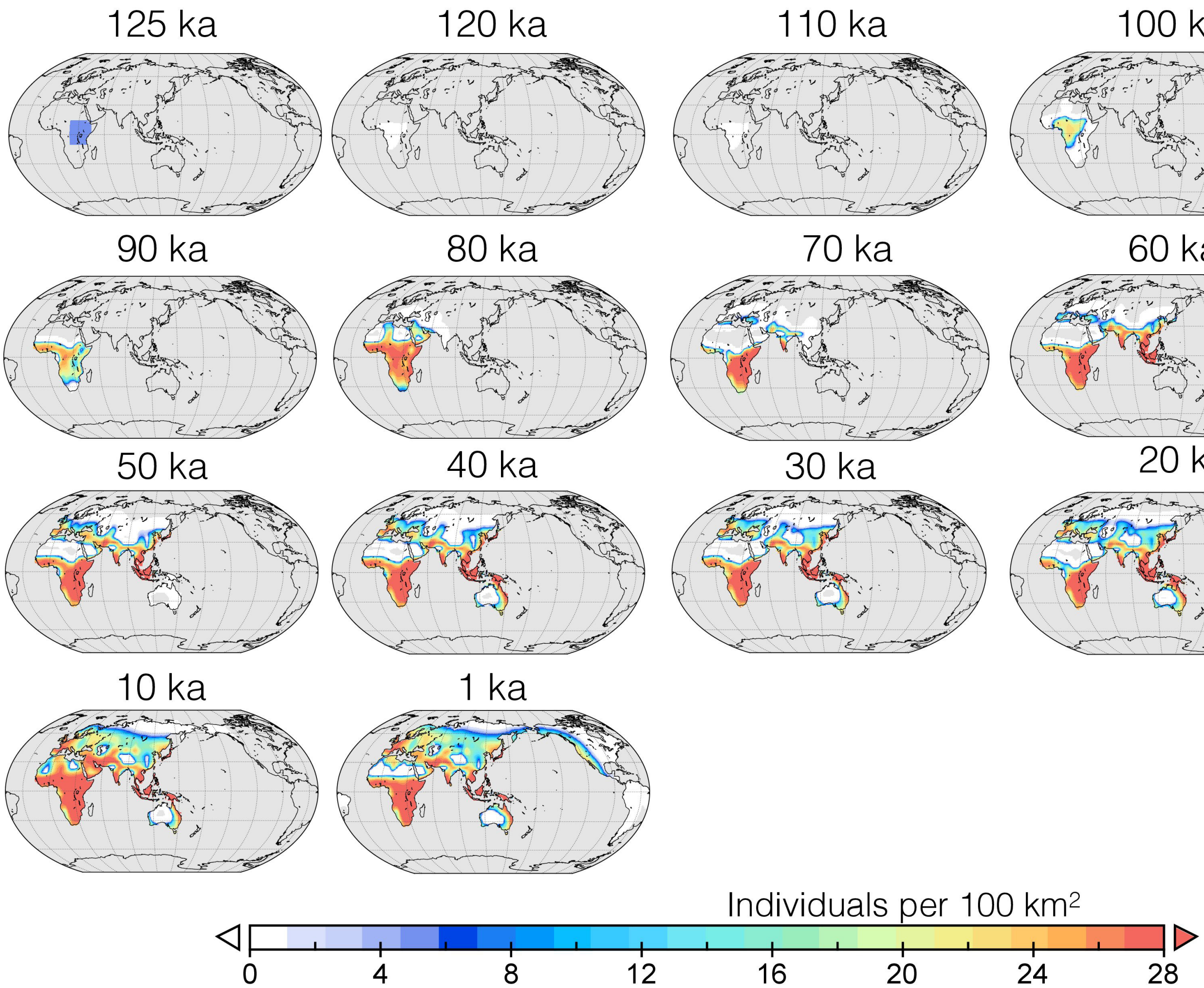
Extended Data Figure 4 | Temperature forcing for HDM. a, First empirical orthogonal function (EOF) of temperature (°C). **b**, The corresponding principal component. First EOF mode captures orbital-scale variability. **c**, Second empirical orthogonal function of temperature (°C). The corresponding principal component is shown in **d**. Second EOF mode captures Heinrich and Dansgaard–Oeschger events. In **b**, the main Marine Isotope Stages (MIS) are indicated with blue shading. In **d**, the blue shading indicates the main Heinrich stadials and the C-events.



Extended Data Figure 5 | Net primary production forcing for HDM. Same as Extended Data Fig. 4, but for primary production (kgC m^{-2} yr^{-1}).



Extended Data Figure 6 | Desert fraction forcing for HDM. Same as Extended Data Fig. 4, but for desert fraction (%).



Extended Data Figure 7 | Late Pleistocene human dispersal. Snapshots of the simulated evolution of human density (individuals per 100 km^2) over the past 125 ka using the parameters of the scenario B (late exit) experiment (see Methods) with full climate (orbital and millennial-scale) and sea level forcing and with human adaptation.

Extended Data Table 1 | Parameter configurations of human dispersal model used in early exit (scenario A) and late exit (scenario B) scenarios

Parameter	Early exit scenario	Late exit scenario
K (main diffusion parameter)	4.25 km^2 $year^{-1}$ to 42 km^2 $year^{-1}$ over 125 ka	2.125 km^2 $year^{-1}$ to 42 km^2 $year^{-1}$ over 125 ka
T_w (Temperature width)	6 °C	7 °C
T_c (critical temperature)	15 °C to -35 °C over 125 ka	17.5 °C to -35 °C over 125 ka
d_w (Desert width)	8%	8%
d_c (critical desertfraction)	45- 70% over 125 ka	45- 80% over 125 ka
N_c (critical net primary productivity)	0.1 kgC/m^2/year	0.1 kgC/m^2/year
N_w (net primary productivity width)	0.3 kgC/m^2/year	0.3 kgC/m^2/year
G (Growth rate)	0.004 $year^{-1}$	0.004 $year^{-1}$
m (Mortality rate)	0.105 $year^{-1}$	0.105 $year^{-1}$
a (nonlinear damping rate)	1.25 10^{-6} $year^{-1}$	1.25 10^{-6} $year^{-1}$
U_p (coastal propagation speed)	0.0687 to 2.4056 km $year^{-1}$ over 125 ka	0.0343 to 2.0619 km $year^{-1}$ over 125 ka

Extended Data Table 2 | Sensitivity experiments conducted with human dispersal model using different climate and dispersal scenarios

Abbreviation	Forcings and experimental set-up
Scenario A (Early exit)	Fully varying climate and sea level conditions 125-0 ka: $T(x,y,t)$, $N(x,y,t)$, $d(x,y,t)$, parameters in Extended Table 1, left column
noDO	Varying orbital-scale climate and sea level conditions 125-0 ka: $T_{orb}(x,y,t)$, $N_{orb}(x,y,t)$, $d_{orb}(x,y,t)$, parameters in Extended Table 1, left column
Scenario B (Late exit)	Fully varying climate and sea level conditions 125-0 ka: $T(x,y,t)$, $N(x,y,t)$, $d(x,y,t)$, parameters in Extended Table 1, right column
105 ka	Perpetual 105 ka climate and sea level conditions for 125,000 simulation years: $T_{orb}(x,y,105\ ka)$, $N_{orb}(x,y,105\ ka)$, $d_{orb}(x,y,105\ ka)$, parameters in Extended Table 1, left column
70 ka	Perpetual 70 ka climate and sea level conditions for 125,000 simulation year: $T_{orb}(x,y,70\ ka)$, $N_{orb}(x,y,70\ ka)$, $d_{orb}(x,y,70\ ka)$, parameters in Extended Table 1, left column
21 ka	Perpetual 21 ka climate and sea level conditions for 125,000 simulation years: $T_{orb}(x,y,21\ ka)$, $N_{orb}(x,y,21\ ka)$, $d_{orb}(x,y,21\ ka)$, parameters in Extended Table 1, left column

doi:10.1038/nature19770

A cross-modal genetic framework for the development and plasticity of sensory pathways

Laura Frangeul[1]*, Gabrielle Pouchelon[1]*†, Ludovic Telley[1]*, Sandrine Lefort[1], Christian Luscher[1,2] & Denis Jabaudon[1,2]

Modality-specific sensory inputs from individual sense organs are processed in parallel in distinct areas of the neocortex. For each sensory modality, input follows a cortico–thalamo–cortical loop in which a 'first-order' exteroceptive thalamic nucleus sends peripheral input to the primary sensory cortex, which projects back to a 'higher order' thalamic nucleus that targets a secondary sensory cortex[1–6]. This conserved circuit motif raises the possibility that shared genetic programs exist across sensory modalities. Here we report that, despite their association with distinct sensory modalities, first-order nuclei in mice are genetically homologous across somatosensory, visual, and auditory pathways, as are higher order nuclei. We further reveal peripheral input-dependent control over the transcriptional identity and connectivity of first-order nuclei by showing that input ablation leads to induction of higher-order-type transcriptional programs and rewiring of higher-order-directed descending cortical input to deprived first-order nuclei. These findings uncover an input-dependent genetic logic for the design and plasticity of sensory pathways, in which conserved developmental programs lead to conserved circuit motifs across sensory modalities.

Somatosensory input reaches the primary and secondary somatosensory cortex via the ventrobasal nucleus (VB) and posteromedial nucleus (Po) of the thalamus, respectively[7–9]. Similarly, visual input reaches the primary and secondary visual cortex via the dorsolateral geniculate nucleus (LG) and pulvinar/latero-posterior nucleus (LP), respectively[10,11] (Fig. 1a). Despite their functional specialization, these parallel sensory pathways have a markedly conserved circuit design[1–6,12,13]. For each of these modalities, sensory input reaches first order (FO) exteroceptive nuclei (that is, the VB and LG), which project to layer 4 (L4) of their corresponding primary cortical area. Higher order (HO) nuclei (that is, the Po and LP) are instead innervated by top-down cortical inputs originating in L5B of the primary cortex[6,11,14] and project to L4 of their secondary cortical area[3] (Fig. 1a, b). This conserved circuit motif prompted us to investigate whether shared genetic programs drive sensory thalamocortical (TC) circuit assembly and plasticity across modalities.

To investigate this possibility, we performed a transcriptional analysis comparing gene expression in mouse VB, Po, LG, and LP neurons at post-natal day 3 (P3) (Fig. 1c), a time at which TC axons are reaching their cortical targets[5]. To characterize the genetic identities of these nuclei, we used dimensionality reduction and unsupervised clustering of samples based on transcriptional signatures[15,16]. This approach revealed that the transcription identities of FO nuclei were closely related across modalities, as were those of HO nuclei (Fig. 1d). To compare the discriminative power of the 'hierarchical order'-based classification (that is, FO versus HO nuclei) with the modality-based classification (that is, somatosensory versus visual nuclei), we trained a support-vector machine (SVM) model (Fig. 1e, Extended Data Fig. 1a, and see Methods) with FO and HO nuclei transcripts, or somatosensory and visual nuclei transcripts. This machine-learning approach revealed that hierarchical order is more discriminative than

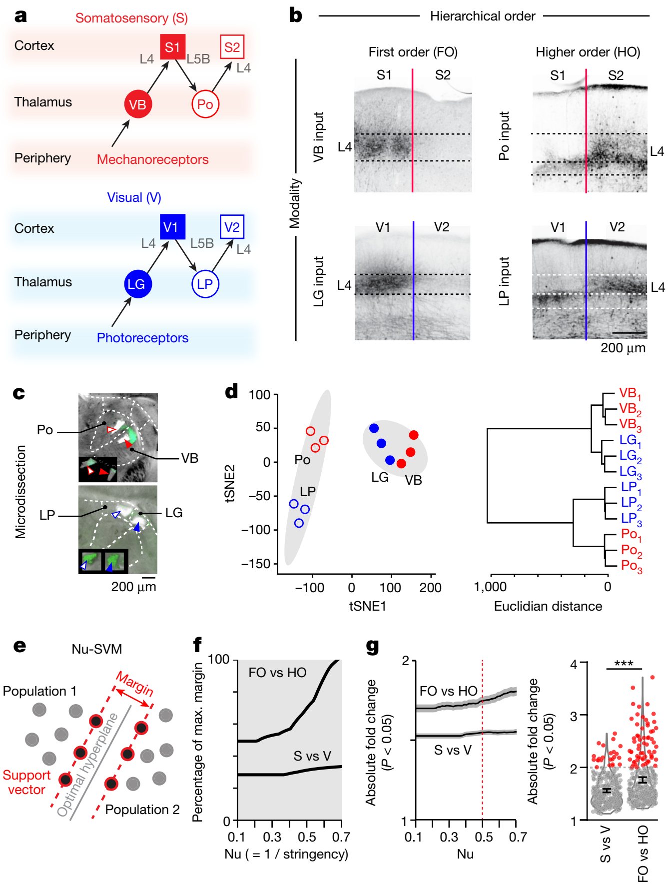


Figure 1 | Hierarchical order is the primary determinant of transcriptional identity in somatosensory and visual thalamic nuclei. a, Schematic representation of somatosensory and visual pathways. **b**, Anterograde labelling from FO and HO nuclei reveals parallel organizations of somatosensory and visual TC pathways. **c**, Illustrative microdissection of thalamic nuclei at P3; acute coronal section. Nuclei were identified by retrograde labelling from the primary somatosensory cortex (S1) or primary visual cortex (V1). S2, secondary somatosensory cortex; V2, secondary visual cortex. **d**, Unbiased clustering delineates FO and HO nuclei. Shaded ellipse represents 85% confidence area around centroid. Circles represent individual samples. tSNE, *t*-distributed stochastic neighbour embedding. **e**, Nu-support vector machine (nu-SVM) analysis identifies the optimal demarcation plane between two populations. High margin values indicate high discriminative power. **f**, FO versus HO delineation is superior to S versus V delineation at all levels of stringency. **g**, Type-specific genes are more differentially expressed between FO and HO nuclei than between somatosensory and visual nuclei. Fold changes represent ratios of expression between significantly differentially expressed genes for each condition. Right panel shows fold changes for nu = 0.5 (red line in left panel). Genes with fold changes >2 are highlighted in red. ***$P < 0.001$; Welch's two-sample *t*-test.

[1]Department of Basic Neurosciences, University of Geneva, 1211 Geneva, Switzerland. [2]Clinic of Neurology, Geneva University Hospital, 1211 Geneva, Switzerland. †Present address: New York University Langone Medical Center, New York, 10016 New York, USA.
*These authors contributed equally to this work.

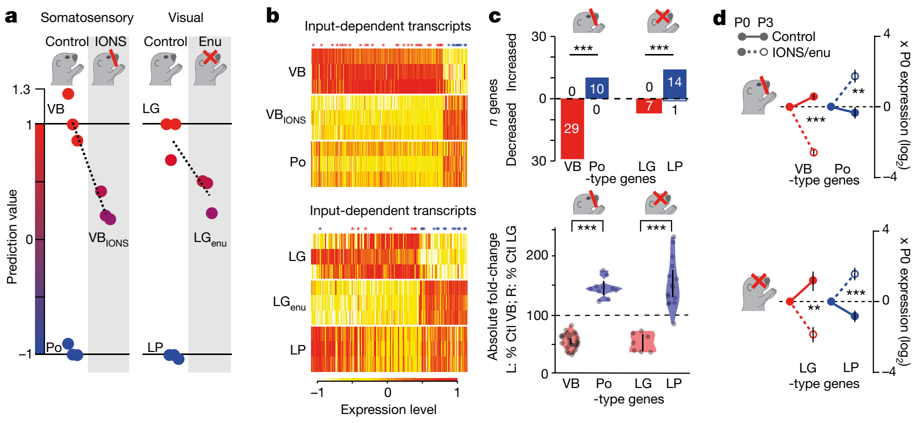


Figure 2 | Peripheral input ablation induces HO-type transcriptional programs in FO nuclei. a, P0 infraorbital nerve section (IONS) or enucleation (enu) leads to decreased expression of corresponding FO-type and induction of HO-type genes in the VB and LG. Linear model using gene expression in corresponding control FO and HO nuclei as training sets. **b**, Expression of peripheral input-dependent genes. Asterisks indicate VB- and Po- or LG- and LP-specific genes from Supplementary Table 2. FO gene expression is decreased and HO gene expression increased in the IONS and enucleation conditions. **c**, Quantification of the data shown in b. L, left; R, right; ctl, control. Top, $***P < 0.001$; Fisher's exact test. Bottom, $***P < 0.001$; Welch's two-sample *t*-test. **d**, Developmental dynamics of FO- and HO-type input-dependent genes in the VB and LG under control and input-ablated conditions. Input ablation prevents the induction of FO-type genes and repression of HO-type genes normally observed between P0 and P3. $**P < 0.01$; $***P < 0.001$; Welch's two-sample *t*-test.

modality at all levels of stringencies examined (Fig. 1f). In agreement with this finding, differentially expressed genes showed significantly higher fold changes when comparing FO and HO nuclei (Fig. 1g, Supplementary Table 1 and Supplementary Note 1). Although SVM margins for hierarchical order were larger than for modality throughout post-natal development, differences were greater at P3 and P10 than at P0, suggesting a dynamic post-natal process (Fig. 1f, Extended Data Fig. 1b, c). Together, these data indicate that hierarchical order, not modality, is the primary determinant of transcriptional identity in somatosensory and visual TC pathways.

The auditory thalamus also consists of an FO nucleus (ventral medial geniculate nucleus, vMG) and HO nucleus (dorsal medial geniculate nucleus, dMG) with corresponding bottom-up and top-down connectivities, albeit with some overlapping circuit features between these two nuclei (Extended Data Fig. 2a, top). For example, in contrast to the mutually exclusive laminar projections of FO and HO in other modalities, the dMG and vMG both project to L1 and do not project to L5A[4,17] (Extended Data Fig. 2a, bottom). Despite this overlap, we hypothesized that the order-based genetic organization identified above might apply, albeit less stringently, to the auditory thalamus. Consistent with this possibility, gene expression analysis revealed that although vMG and dMG were transcriptionally related at P3, FO- and HO-nuclei-specific genes were enriched in the vMG and dMG, respectively (Extended Data Fig. 2b, c, Supplementary Table 2). Conversely, classification using all vMG- and dMG-specific transcripts as training sets showed a corresponding hierarchical order-based distribution of somatosensory and visual nuclei (Extended Data Fig. 2d). Finally, *in situ* hybridization detecting FO- and HO-specific transcripts showed consistent expression across modalities in somatosensory, visual, and auditory nuclei (Extended Data Fig. 2e, f). Together, these data indicate that a hierarchical order-based genetic parcellation of thalamic nuclei applies across sensory modalities.

As FO nuclei receive input from their respective peripheral pathways, it is possible that sensory inputs drive the acquisition of FO transcriptional identity. To investigate this possibility, we ablated peripheral input to the VB by performing a neonatal infraorbital nerve section (IONS)[8], or ablated retinal input to the LG by performing a neonatal bilateral enucleation[14]. These two manipulations represent selective and equivalent ablations of input to somatosensory and visual pathways, respectively. Consistent with a periphery-dependent control of FO neuron differentiation, IONS and enucleation had the same effects on VB and LG gene expression at P3 (Fig. 2). Prediction of transcriptional identity using control conditions (that is, VB and Po transcripts, and LG and LP transcripts) as training sets revealed an FO to HO shift in the identity of deprived VB and LG (Fig. 2a). Both procedures decreased the expression of their corresponding FO-type genes and led to the emergence of HO-type transcript expression in the VB and LG (Fig. 2b, c, Extended Data Fig. 3a–c, Supplementary Table 3, Supplementary Note 2). These findings suggest that an input-dependent repression of HO transcripts may normally occur in exteroceptive FO nuclei. Supporting this possibility, between P0 and P3 we observed an induction of FO-type genes and repression of HO-type genes in control conditions; this transcriptional balance was impaired by input ablation (Fig. 2d, Extended Data Fig. 3d). Notably, in contrast to exteroceptive FO nuclei, the transcriptional identity of HO nuclei was less affected by peripheral input ablation, consistent with their largely cortically driven inputs[1,2,8,9,11] (Extended Data Fig. 4). Together, these experiments reveal that acquisition of final FO nucleus identity is a process engaged by sensory stimulation, occurring through post-natal induction of FO-type genes and repression of HO-type genes.

Expression of HO-type transcriptional programs in input-ablated FO nuclei could be associated with the abnormal acquisition of HO-directed L5B synaptic input by deprived FO neurons[6,11,14,18]. As a first approach to investigate this possibility, we genetically marked[18,19] the presynaptic terminals of L5B neurons. In contrast to control conditions, numerous L5B presynaptic terminals were present in the input-ablated LG, confirming anatomical cross-hierarchical rewiring in the absence of peripheral input[14] (Extended Data Fig. 5a). To demonstrate the acquisition of functional L5B input by deprived LG neurons, we optogenetically stimulated L5B axons following injection of a conditional channelrhodopsin-2 (ChR2)-expressing recombinant adeno-associated virus into the visual cortex of *Rbp4*-Cre mice, which express Cre recombinase in L5 neurons[19] (Fig. 3a, Extended Data Fig. 5b). Although LG neurons essentially did not receive L5B input under normal conditions (22 out of 23 recorded neurons), connection probability markedly increased in enucleated mice, reaching levels comparable to those found in the normal LP (Fig. 3b, Extended Data Fig. 5c). Therefore, loss of peripheral input leads to acquisition of HO-directed descending cortical input by input-ablated FO neurons, which matches the change in their transcriptional identities (Fig. 3c).

Our findings reveal a developmental genetic framework for the conserved design of TC circuits across sensory modalities. These conserved developmental programs constitute a parsimonious strategy to set up core sensory circuit motifs in the immature brain. Eventually, once these motifs are set up, the specificities of each modality can

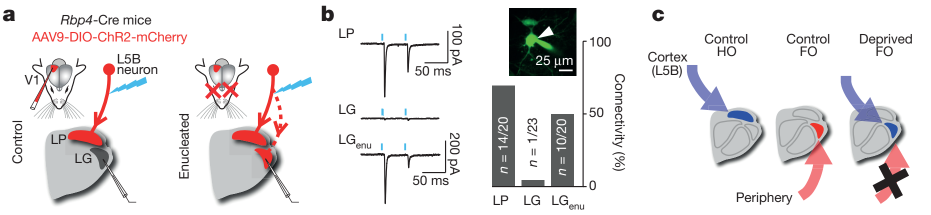


Figure 3 | Enucleation leads to acquisition of LP-directed descending L5B input by deprived LG neurons. a, Schematic representation of the experimental setting and hypothesis. **b**, Optogenetic stimulation of the axons of V1 L5B neurons elicits postsynaptic responses in LG neurons following enucleation. Inset, recorded neuron in the LG filled with a fluorescent dye. Arrowhead indicates tip of the patch pipette. Percentage of connectivity of LG versus LG_{enu}: $P < 0.001$; LG_{enu} versus LP: not significant; χ^2 statistics on contingency table. **c**, Summary of the findings.

then emerge in an input-dependent manner as development ends and sensory pathways become operationally mature.

The results of our input ablation experiments support the idea that HO-type genetic identity is a ground-state feature[20,21], perhaps imparted by internal, top-down inputs, from which FO-type identity emerges in a peripheral input-dependent manner. FO-type TC neurons may thus have emerged from ancestral HO-type neurons through selection for specific metabolic[22,23], electrophysiological[24], and precise connectivity features required to convey signals generated by high-resolution body receptors. Furthermore, the transcriptional homologies identified here may relate to synaesthesia and to the ability of sensory pathways to extensively rewire across modalities following lesions at early developmental stages[25]. The input-dependent genetic logic for sensory circuit construction presented here thus provides a unifying mechanism for the evolutionary emergence and cross-modal plasticity of sensory pathways.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 6 June; accepted 16 August 2016.
Published online 26 September 2016.

1. Guillery, R. W. Anatomical evidence concerning the role of the thalamus in corticocortical communication: a brief review. *J. Anat.* **187,** 583–592 (1995).
2. Sherman, S. M. S. Thalamocortical interactions. *Curr. Opin. Neurobiol.* **22,** 575–579 (2012).
3. Herkenham, M. Laminar organization of thalamic projections to the rat neocortex. *Science* **207,** 532–535 (1980).
4. Smith, P. H., Uhlrich, D. J., Manning, K. A. & Banks, M. I. Thalamocortical projections to rat auditory cortex from the ventral and dorsal divisions of the medial geniculate nucleus. *J. Comp. Neurol.* **520,** 34–51 (2012).
5. Clascá, F., Rubio-Garrido, P. & Jabaudon, D. Unveiling the diversity of thalamocortical neuron subtypes.*Eur. J. Neurosci.* **35,** 1524–1532 (2012).
6. Theyel, B. B., Llano, D. A. & Sherman, S. M. The corticothalamocortical circuit drives higher-order cortex in the mouse. *Nat. Neurosci.* **13,** 84–88 (2010).
7. Pouchelon, G. *et al.* Modality-specific thalamocortical inputs instruct the identity of postsynaptic L4 neurons. *Nature* **511,** 471–474 (2014).
8. Frangeul, L. *et al.* Specific activation of the paralemniscal pathway during nociception. *Eur. J. Neurosci.* **39,** 1455–1464 (2014).
9. Viaene, A. N., Petrof, I. & Sherman, S. M. Properties of the thalamic projection from the posterior medial nucleus to primary and secondary somatosensory cortices in the mouse. *Proc. Natl Acad. Sci. USA* **108,** 18156–18161 (2011).
10. Chou, S.-J. *et al.* Geniculocortical input drives genetic distinctions between primary and higher-order visual areas. *Science* **340,** 1239–1242 (2013).
11. Wang, S., Eisenback, M. A. & Bickford, M. E. Relative distribution of synapses in the pulvinar nucleus of the cat: implications regarding the "driver/modulator" theory of thalamic function. *J. Comp. Neurol.* **454,** 482–494 (2002).
12. Reichova, I. & Sherman, S. M. Somatosensory corticothalamic projections: distinguishing drivers from modulators. *J. Neurophysiol.* **92,** 2185–2197 (2004).
13. Lee, C. C. & Sherman, S. M. Synaptic properties of thalamic and intracortical inputs to layer 4 of the first- and higher-order cortical areas in the auditory and somatosensory systems. *J. Neurophysiol.* **100,** 317–326 (2008).
14. Grant, E., Hoerder-Suabedissen, A. & Molnár, Z. The regulation of corticofugal fiber targeting by retinal inputs. *Cereb. Cortex* **26,** 1336–1348 (2016).
15. Telley, L. *et al.* Sequential transcriptional waves direct the differentiation of newborn neurons in the mouse neocortex. *Science* **351,** 1443–1446 (2016).
16. Macosko, E. Z. *et al.* Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161,** 1202–1214 (2015).
17. Lu, E., Llano, D. A. & Sherman, S. M. Different distributions of calbindin and calretinin immunostaining across the medial and dorsal divisions of the mouse medial geniculate body. *Hear. Res.* **257,** 16–23 (2009).
18. Shanks, J. A. *et al.* Corticothalamic axons are essential for retinal ganglion cell axon targeting to the mouse dorsal lateral geniculate nucleus. *J. Neurosci.* **36,** 5252–5263 (2016).
19. Gong, S. *et al.* Targeting Cre recombinase to specific neuron populations with bacterial artificial chromosome constructs. *J. Neurosci.* **27,** 9817–9823 (2007).
20. Butler, A. B. Evolution of the thalamus: a morphological and functional review. *Thalamus Relat. Syst.* **4,** 1–24 (2008).
21. Bishop, G. H. The relation between nerve fiber size and sensory modality: phylogenetic implications of the afferent innervation of cortex. *J. Nerv. Ment. Dis.* **128,** 89–114 (1959).
22. Wong-Riley, M. T. & Welt, C. Histochemical changes in cytochrome oxidase of cortical barrels after vibrissal removal in neonatal and adult mice. *Proc. Natl Acad. Sci. USA* **77,** 2333–2337 (1980).
23. Wong-Riley, M. Changes in the visual system of monocularly sutured or enucleated cats demonstrable with cytochrome oxidase histochemistry. *Brain Res.* **171,** 11–28 (1979).
24. Yamashita, T. *et al.* Membrane potential dynamics of neocortical projection neurons driving target-specific signals. *Neuron* **80,** 1477–1490 (2013).
25. Bavelier, D. & Neville, H. J. Cross-modal plasticity: where and how? *Nat. Rev. Neurosci.* **3,** 443–452 (2002).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank A. Benoit, S. Binvignat, M. Lanzillo and members of the Genomics Platform of the University of Geneva for technical assistance, A. Holtmaat and A. Carleton for the gift of the transgenic mouse lines. We thank J. Prados for help in the bioinformatics analysis. We thank E. Azim, A. Holtmaat, M. Scanziani and S. Tole for helpful comments on the manuscript. Work in the Jabaudon laboratory is supported by the Swiss National Science Foundation (SNF) (PP00P3_123447), the Leenaards Foundation, the Synapsis Foundation and the NARSAD Foundation.

Author Contributions G.P., L.F. and D.J. conceived the project; G.P., L.F. and S.L. performed the experiments; L.T., G.P., L.F., S.L. and D.J. performed analyses; D.J. and L.F. wrote the manuscript with help of all authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.J. (denis.jabaudon@unige.ch).

Reviewer Information *Nature* thanks S. Nelson, F. Polleux and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Mice. C57BL/6 male and female P0, P3, P10 pups and adult mice were used. Randomization was not applicable. Transgenic mice consist of *Gt(ROSA)26Sor*$^{tm34.1(CAG-Syp/tdTomato)Hze}$ ('Syp-tdT', Jackson Laboratories, stock number 012570) and Tg(*Rbp4*-Cre)KL100Gsat/Mmcd (denoted as *Rbp4*-Cre, GENSAT RP24-285K21) backgrounds. Experiments were carried out with permission of the Geneva cantonal authorities.

Histology. *In situ* hybridization on slides was performed according to methods described previously[7]. For antisense probe synthesis, total complementary DNA was amplified by PCR with primers designed for the specific messenger RNA sequence of *Id2*, *Cdkn1c*, *Glra3*, *Nxph1* and *Tcf7l2*. T7 or Sp6 promoter sequence was added to the reverse primer sequence. Digoxigenin (DIG)-labelled antisense RNA probes were obtained after *in vitro* transcription of the resulting PCR product (Roche kit). Briefly, hybridization was carried out overnight at 60 °C with the DIG-labelled RNA probes. After hybridization, sections were washed and incubated with alkaline phosphatase-conjugated anti-DIG antibody (1:2,000, Roche, 11093274910) overnight at 4 °C. Sections were then washed and the revelation procedure was carried out at room temperature in a solution containing NBT (nitro-blue tetrazolium chloride) and BCIP (5-bromo-4-chloro-39-indoly phosphate p-toluidine salt) (Roche, 11681451001). After revelation, sections were washed, post-fixed for 30 min in 4% PFA and mounted with Fluoromount (Sigma, F4680).

Anterograde/retrograde labelling. For retrograde labelling from the cortex (Fig. 1b, c, Extended Data Fig. 2a), anaesthetized P0 pups were placed on a stereotaxic apparatus and thalamic neurons were retrogradely labelled via 55-nl cortical injections of Green IX Retrobeads (Lumafluor, Inc., G180). Pup handling was performed as described in ref. 26. To label TC axons (Fig. 1b, Extended Data Fig. 2a), virus (AAV1-hSynap-eGFP-WPRE-bGH) was stereotaxically injected into the VB, Po, LG, LP, vMG and dMG of adult mice. For VB injections, coordinates (in mm) were: AP, −2.1 and ML, ± 1.6 from the bregma; DV, −3.0 from the pial surface. For Po injections, coordinates were: AP, −2.1; and ML, ± 1.3 from the bregma; DV, −3.0 from the pial surface. For LG injections, coordinates were: AP, −2.1 and ML, ± 1.9 from the bregma; DV, −2.7 from the pial surface. For LP injections, coordinates were: AP, −2.1; and ML, ± 1.3 from the bregma; DV, −2.5 from the pial surface. For vMG injections, coordinates were: AP, −3.3 and ML, ± 2.2 from the bregma; DV, −3.5 from the pial surface. For dMG injections, coordinates were: AP, −3.1 and ML, ± 2 from the bregma; DV, −2.7 from the pial surface. Brains were perfusion-fixed after 14 days and serially sectioned on a vibrating microtome at 50 μm. To label cortico-thalamic axons in *Rbp4*-Cre mice (Fig. 3), virus (AAV9-EF1α-DIO-ChR2-mCherry) was stereotaxically injected into the visual cortex of P0 pups. Pup handling was performed as described in ref. 26.

ION section and enucleation. Infraorbital nerve section (IONS) and enucleation were performed on P0 pups. Animals were deeply anaesthetized on ice. The right ION was sectioned as previously described[8]. For enucleation, a small incision was made in the eyelid with a scalpel and the eye was separated from the optic nerve with microscissors in order to be removed from the orbit with forceps. Pups were briefly warmed on a heating pad and were returned to their mother.

Tissue microdissection, RNA amplification, and microarray hybridization. Fresh coronal brain sections (140 μm) were cut on a vibrating microtome and thalamic nuclei (retrogradely-labelled for VB, Po, LG, LP, vMG and dMG P0 and P3 collections) were visually identified and microdissected using a stereomicroscope (Leica, M165FC) in ice-cold oxygenated artificial cerebrospinal fluid under RNase-free conditions. Samples were collected at P0, P3 and P10 (P3 only for vMG, dMG, IONS and enu), yielding a total of 9 VB, 9 Po, 9 LG, 9 LP, 3 vMG, 3 dMG, 3 VB_{IONS}, 3 Po_{IONS}, 3 LG_{enu} and 3 LP_{enu} samples, which were stored in RNA later at −80 °C. RNA was extracted using an RNeasy kit (Qiagen) and two-cycle amplification and labelling was performed according to Affymetrix protocols using Superscript cDNA synthesis kit (Invitrogen), MEGAscript T7 kit and MessageAmp IIaRNA amplification kit (Ambion). Experiments were performed blindly. Labelled cRNA was fragmented and hybridized to Affymetrix Mouse Genome 430 2.0 Array. GeneChips were incubated at 45 °C for 16 h with biotin-labelled cRNA probes, and then washed and stained using a streptavidin–phycoerythrin conjugate with antibody amplification as described in Affymetrix protocol, using Affymetrix GeneChip Fluidics Station 450. GeneChips were scanned on a GCS3000 scanner (Affymetrix). Some of these arrays were used for identification of interneuron-specific genes in ref. 27.

Data analysis. Microarray CEL files were read and normalized using 'affy' and 'gcrma' R packages and transformed in $\log_2$. All probesets with an expression value $>\log_2 10$ in at least 10% of all samples were included in the analysis (23,767 probesets)[15]. *t*-distributed stochastic neighbour embedding (Fig. 1d, Extended Data Fig. 2b) (Rtsne, R package, 'perplexity parameter' = 3) returns a two-dimensional embedding of samples. Samples with similar expression of genes and therefore similar principal components loadings, are most likely to localize near each other in the embedding[15,16]. Hierarchical clustering was performed using Euclidian distance metrics. To compare the discriminative power of the 'hierarchical order'-based classification (that is, FO versus HO nuclei) with the modality-based classification (that is, somatosensory versus visual nuclei) (Fig. 1e–g, Extended Data Fig. 1), we trained 2 linear nu-support vector machine (nu-SVM) classification models[28–30], (P3 in Fig. 1 and P0; P10 in Extended Data Fig. 1). Nu corresponds to the degrees of freedom of the SVM model, and thus inversely correlates with stringency. We determined the maximal margin of separation between the two populations (that is, HO versus FO, or somatosensory versus visual), which indicates how distinct these two populations are. Because the 'nu' parameter controls the stringency of the model[28], we confirmed the results using a range of nu values between 0.1 and 0.7; nu = 0.5 was used for further analyses. Cross-validation (leave-one-out) was performed for all models trained to avoid overfitting[28]. Differentially expressed genes were identified based on their weight in the SVM prediction model (Fig. 1g, Extended Data Fig. 2), looking for outliers in its linear coefficients using a *z*-test and a false discovery rate (FDR)-corrected P value < 0.05. This identified 379 hierarchical order-specific probesets (142 FO, 237 HO) that were used to generate the heat map in Extended Data Fig. 2. Modality-specific transcripts originated from 332 probesets (155 s, 177 V). Differentially expressed genes presented in these heat maps were identified as described above (Fig. 2b, Extended Data Fig. 4b). VB versus VB_{IONS}, Po versus Po_{IONS}, LG versus LG_{enu}, LP versus LP_{enu} comparisons were performed, which identified nucleus-specific input-dependent probesets (VB versus VB_{IONS}: 42 increased, 231 decreased; Po versus Po_{IONS}: 88 increased, 127 decreased; LG versus LG_{enu}: 55 increased, 123 decreased; LP versus LP_{enu}: 158 increased, 91 decreased). Nucleus-specific genes were identified as described above (Fig. 2c, Extended Data Fig. 3, 4). VB versus Po, LG versus LP comparisons were performed, which identified nucleus-specific probesets (180 VB, 237 Po, 110 LG, 237 LP). These gene lists were intersected with the corresponding input-dependent gene list (for example, the 180 VB genes with the 273 VB_{IONS} genes, which yielded 29 common transcripts) to identify nucleus-specific, input dependent genes. The SVM model can predict the transcriptional relationship of new samples with regard to samples used as training sets[28–30] (Fig. 2a, Extended Data Fig. 2d, 3d, 4a). Using vMG and dMG transcripts as training sets, we thus predicted the transcriptional relationship of VB, Po, LG, LP nuclei with these two auditory nuclei (Extended Data Fig. 2d). Similarly VB and Po transcripts were used as training sets to predict the identity of VB_{IONS} (Fig. 2a); LG and LP were used as training sets to predict the identity of LG_{enu} (Fig. 2a); VB P0 and VB P3 were used as training sets to predict the identity of VB_{IONS} (Extended Data Fig. 3d); LG P0 and LG P3 were used as training sets to predict the identity of LG_{enu} (Extended Data Fig. 3d); VB and Po were used as training sets to predict the identity Po_{IONS} (Extended Data Fig. 4a); and LG and LP were used as training sets to predict the identity of LP_{enu} (Extended Data Fig. 4a).

Code availability. Analyses were performed using dedicated R packages including Affy, Gcrma, Hclust, Rtsne, SVM(e1071) which are freely available at https://www.bioconductor.org and https://cran.r-project.org.

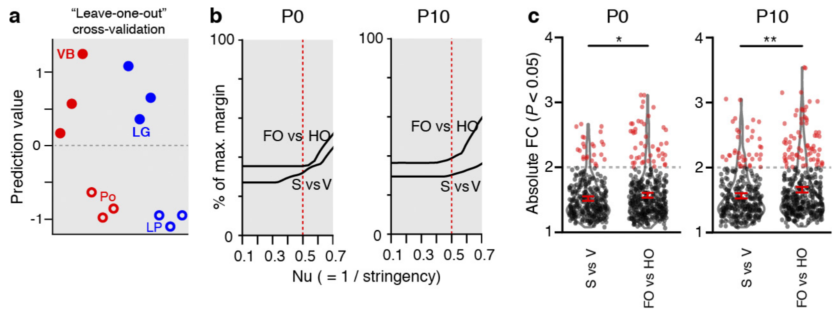
Gene ontologies. Gene ontologies were determined using the Genego portal (https://portal.genego.com/cgi/data_manager.cgi).

Electrophysiology. Two weeks after AAV virus injection, mice (3 enucleated and 4 controls) were collected for electrophysiological recordings. P15–P18 mice were deeply anaesthetized with isoflurane and then decapitated. Coronal 230-μm-thick brain slices were cut on a vibrating microtome (Leica VT1200S) in standard chilled artificial cerebrospinal fluid containing (in mM): 119 NaCl, 2.5 KCl, 1.3 $MgCl_2$, 2.5 $CaCl_2$, 1 Na_2HPO_4, 26.2 $NaHCO_3$ and 11 glucose, bubbled with 95% CO_2 and 5% O_2. Slices were then transferred to a chamber filled with oxycarbonated artificial cerebrospinal fluid at 35 °C for 15 min and then kept at room temperature until use. Electrophysiological recordings were carried out at 35 °C and thalamic neurons (either from LG or LP) were visualized with a two-photon microscope (Femtomics, Hungary) coupled to a 40/0.8 NA objective (Olympus). Whole-cell recordings were obtained with a Multiclamp 700B amplifier (Axon Instruments) and patch-pipettes (3–5 MΩ) filled with (in mM): 130 K-gluconate, 10 Na_2-phosphocreatine, 4 $MgCl_2$, 3.4 Na_2ATP, 0.1 Na_3GTP, 1.1 EGTA, 5 HEPES (pH 7.3, 289 mOsm) complemented with 40 μM CF488A hydrazide (Biotium) for live imaging and 2 mg ml^{-1} Biocytin (Biotium) for post-hoc anatomical staining. Electrophysiological data were amplified, Bessel-filtered at 4 kHz and sampled at 10 kHz (National Instrument). Recorded neurons were clamped at −70 mV and the liquid junction potential was not corrected. Excitatory photosynaptic currents were evoked by 3-ms-long light flashes from a 473 nm solid-state laser delivered through a fibre optic (Thorlabs) directed onto the slice. GABA-mediated inhibitory synaptic transmission was blocked throughout experiments by washing

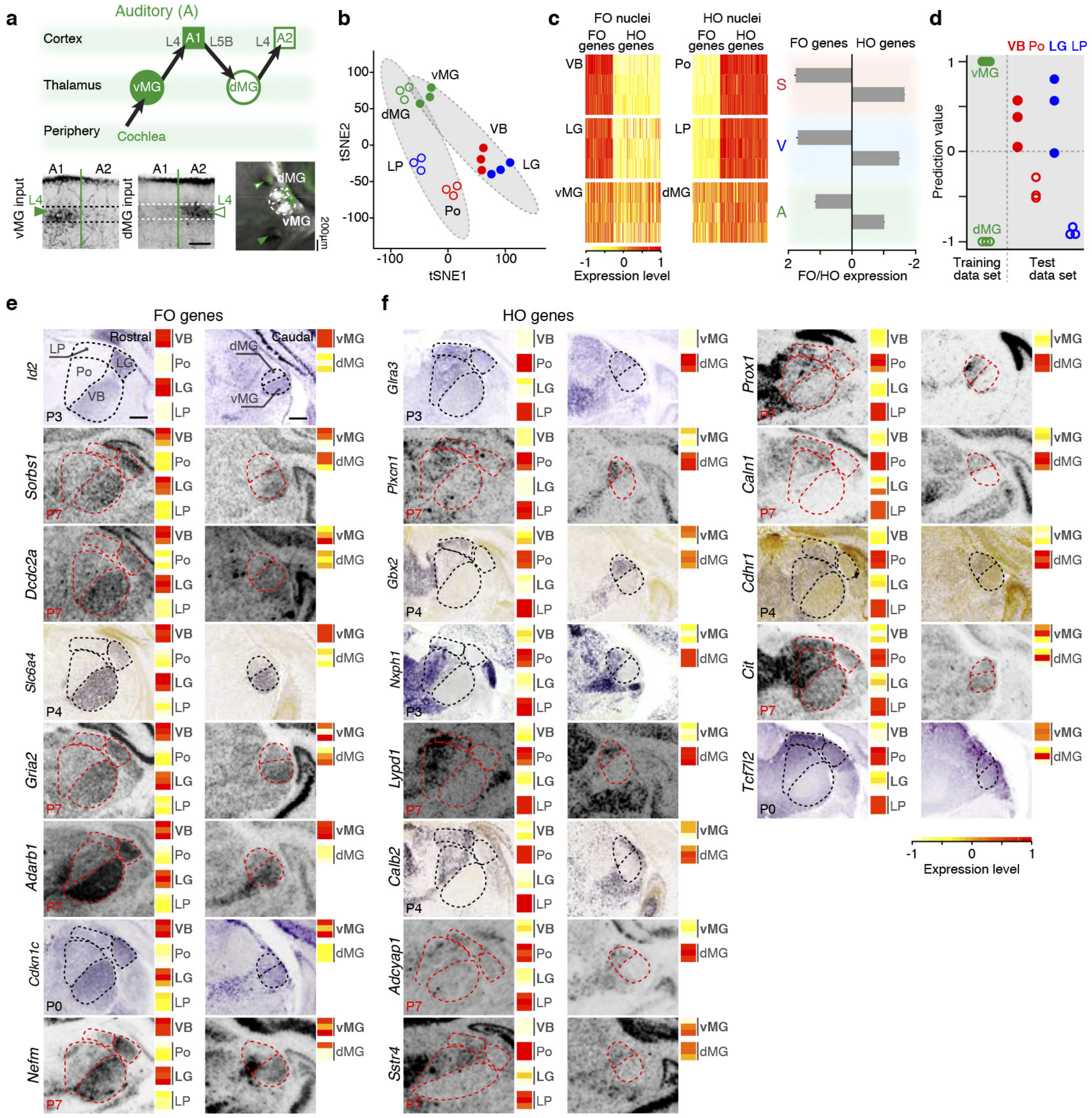
picrotoxin (100 μM, Tocris). Electrophysiological data analysis and statistics were performed using custom routines in IgorPro (Wavemetrics). Data analyses are expressed as mean ± s.e.m. Mean traces were obtained by averaging 30 sweeps. Statistically significant differences in connectivity ($P < 0.05$) were computed by χ^2 statistics on contingency table. Statistically significant differences ($P < 0.05$) for the mean excitatory synaptic amplitude were assessed by performing a non-parametric Kruskall–Wallis test followed by a post-hoc Dunn–Holland–Wolfe test for pairwise comparison.

Statistics. No statistics were used to determine group sample size; however, sample sizes were similar to those used in previous publications from our group and others. Transcriptional analyses were performed blindly. For enucleation and infraorbital nerve sections, the investigator was not blinded. Tests were performed assuming equal variances except when indicated. Values are shown as mean ± s.e.m. throughout the manuscript. *n* values refer to gene numbers unless specified otherwise.

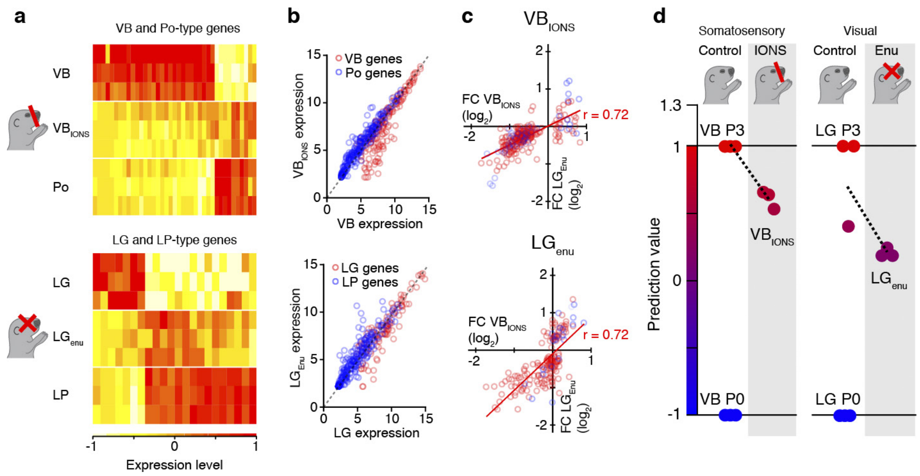
26. De la Rossa, A. & Jabaudon, D. *In vivo* rapid gene delivery into postmitotic neocortical neurons using iontoporation. *Nat. Protocols* **10,** 25–32 (2015).
27. Golding, B. *et al.* Retinal input directs the recruitment of inhibitory interneurons into thalamic visual circuits. *Neuron* **81,** 1057–1069 (2014).
28. Scholkopf, B., Smola, A. J., Williamson, R. C. & Bartlett, P. L. New support vector algorithms. *Neural Comput.* **12,** 1207–1245 (2000).
29. Noble, W. S. What is a support vector machine? *Nat. Biotechnol.* **24,** 1565–1567 (2006).
30. Guyon, I., Weston, J., Barnhill, S. & Vapnik, V. Gene selection for cancer classification using support vector machines. *Mach. Learn.* **46,** 389–422 (2002).
31. Magdaleno, S. *et al.* BGEM: an in situ hybridization database of gene expression in the embryonic and adult mouse nervous system. *PLoS Biol.* **4,** e86 (2006).
32. Bian, W.-J., Miao, W.-Y., He, S.-J., Qiu, Z. & Yu, X. Coordinated spine pruning and maturation mediated by inter-spine competition for cadherin/catenin complexes. *Cell* **162,** 808–822 (2015).



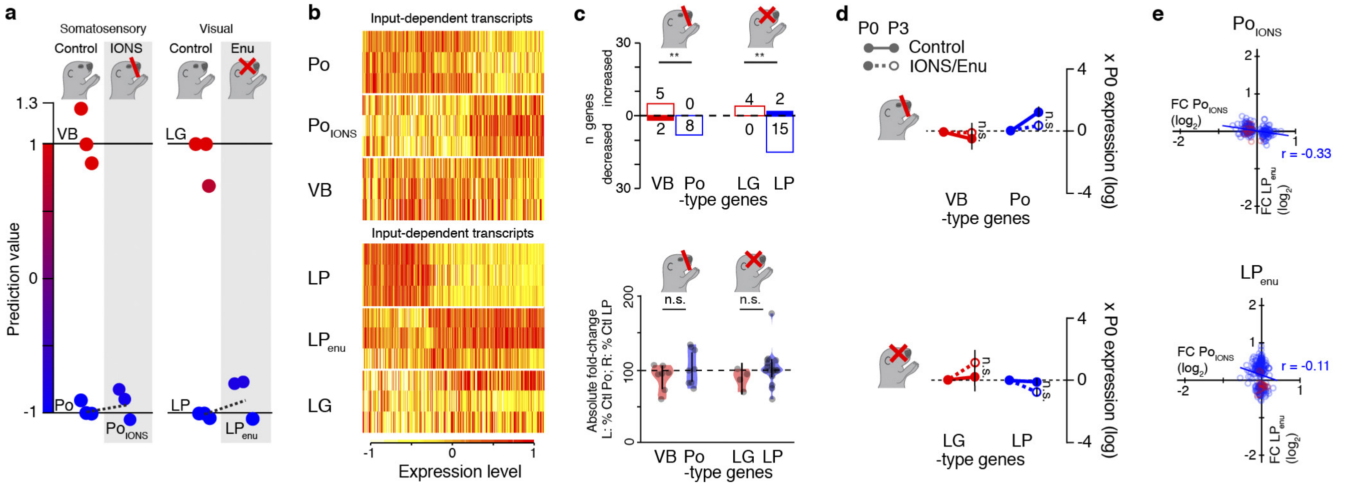
Extended Data Figure 1 | Hierarchical order is the primary determinant of transcriptional identity in somatosensory and visual thalamic nuclei at P0 and P10. a, 'Leave-one-out' cross-validation analysis confirms the robustness of the support vector machine model at P3. See Methods for details. **b**, FO versus HO delineation is superior to the S versus V delineation at all levels of stringency. **c**, Type-specific genes are more differentially expressed between FO and HO nuclei than between S and V nuclei. Fold changes represent ratios of expression between significantly differentially expressed genes for each condition. *$P < 0.05$; **$P < 0.01$; Welch's two-sample *t*-test.



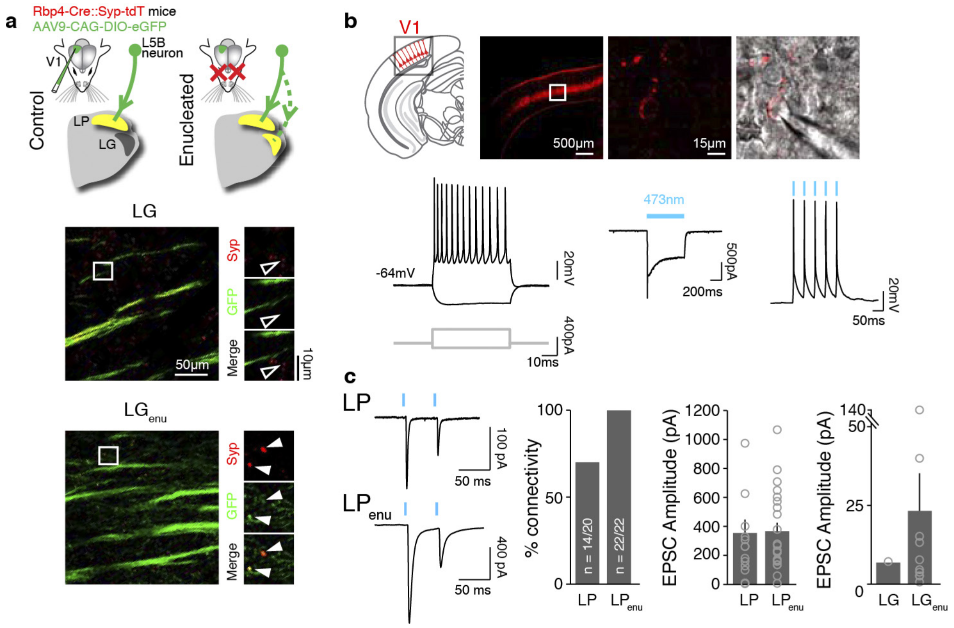
Extended Data Figure 2 | Hierarchical order-based transcriptional logic applies across sensory modalities. a, Top, schematic representation of auditory pathways. Bottom left, anterograde labelling from vMG and dMG. Note that in contrast to somatosensory and visual modalities, both nuclei project to L1 and avoid L5A. Bottom right, Illustrative microdissection, acute coronal section. Nuclei were identified by retrograde labelling from A1. **b**, Unbiased clustering delineates FO and HO nuclei. Shaded ellipse represents 85% confidence area around centroid. Circles represent individual samples. **c**, Expression of the FO- and HO-specific transcripts ('FO genes', 'HO genes') in distinct sensory nuclei. Error bars in the right panel indicate s.e.m. **d**, Unbiased classification using vMG- and dMG-specific transcripts as training sets showed a corresponding hierarchical order-based distribution of somatosensory and visual nuclei. **e**, **f**, *In situ* hybridization on coronal sections showing 8 FO (**e**) and 13 HO-specific transcripts (**f**) and their corresponding level of expression in microarray data. P4 *in situ* hybridization of *Slc6a4*, *Gbx2*, *Calb2* and *Cdhr1* are from the Allen Brain Atlas. P7 *in situ* hybridization of *Sorbs1*, *Dcdc2a*, *Gria2*, *Adarb1*, *Nefm*, *Plxcn1*, *Lypd1*, *Adcyap1*, *Sstr4*, *Prox1*, *Caln1* and *Cit* are from the St Jude Brain Gene Expression Map (BGEM, hosted at http://www.gensat.org)[31]. *In situ* hybridization for *Id2*, *Cdkn1c*, *Glra3*, *Nxph1* and *Tcf7l2* are not available within these databases and were performed based on their high fold-change in gene expression in FO versus HO. Scale bar, 200 μm.



Extended Data Figure 3 | Acquisition of final FO neuron identity is a periphery-dependent process. a, Expression of VB and LG peripheral input-dependent genes (highlighted with an asterisk in Fig. 2). **b**, Top, IONS leads to decreased expression of VB-type genes and increased expression of Po-type genes. Bottom, Enucleation (enu) leads to decreased expression of LG-type genes and increased expression of LP-type genes. **c**, IONS and enucleation affect overlapping sets of genes. Top, genes whose expression is modified by IONS are congruently affected by enucleation. Bottom, genes whose expression is modified by enucleation are congruently affected by IONS. Expression of input-dependent genes in both condition is thus highly correlated (VB_{IONS}: $P < 0.0001$; LG_{enu} $P < 0.0001$). r, regression coefficient. **d**, P0 IONS or enucleation prevents the induction of transcriptional programs normally observed between P0 and P3. Linear model using gene expression in control VB at P0 and P3 or control LG at P0 and P3 as training sets.



Extended Data Figure 4 | HO nucleus identity is largely independent of peripheral input. a, P0 IONS or enucleation does not detectably affect gene expression in the Po and LP. Linear model using gene expression in corresponding control FO and HO nuclei as training sets. **b**, Expression of peripheral input-dependent genes in the Po and LP. **c**, Quantification of the data shown in **b**. L, left; R, right; Ctl, control; n.s., not significant. Top, $^{**}P < 0.01$ (Fisher's exact test); bottom, $P =$ n.s. (Welch's two-sample t-test). **d**, Input ablation does not affect HO developmental dynamics. $P =$ n.s. (Welch's two-sample t-test). **e**, IONS and enucleation affect distinct sets of genes. Top, genes with expression that is modified by IONS are not affected by enucleation. Bottom, genes with expression that is modified by enucleation are not affected by IONS. Expression of input-dependent genes in both condition is thus not correlated (top: Po_{IONS} $r = -0.33$, $P < 0.0001$; bottom: LP_{enu} $r = -0.11$, $P =$ not significant).



Extended Data Figure 5 | Probing L5B connectivity in *Rbp4*-Cre mice. **a**, Top, schematic representation of the experimental setting and hypothesis. Centre, bottom, the presynaptic terminals of L5B neurons were revealed by crossing *Rbp4*-Cre mice, in which Cre recombinase is selectively expressed by L5B neurons[19], with floxed mutants expressing a tomato-red-tagged version of the presynaptic protein synaptophysin (Syp)[32]. In contrast to normal LG, L5B presynaptic terminals are present in LG following enucleation. **b**, Top, *Rbp4*-Cre$^+$ L5B neurons express mCherry. Bottom left, Burst-firing of L5B mCherry$^+$ neurons. Bottom right: example of photocurrents recorded in voltage-clamp during continuous blue-light stimulation and action potentials induced by blue light stimulation. **c**, Optogenetic stimulation of primary visual cortex L5B input elicits postsynaptic responses in LP control and LP_{enu}. Percentage of connectivity of LP versus LP_{enu}: not significant; χ^2 statistics on contingency table. EPSC amplitude (pA): LP ($n = 14$ out of 20), 353.1 ± 94.6; LP_{enu} ($n = 22$ out of 22), 365.5 ± 57.9; LG ($n = 1$ out of 23), 8.6; LG_{enu} ($n = 10$ out of 20), 24.5 ± 13.7. Values are shown as mean $\pm$ s.e.m.

LETTER

doi:10.1038/nature19766

Autocrine BDNF–TrkB signalling within a single dendritic spine

Stephen C. Harward[1]*, Nathan G. Hedrick[1]*†, Charles E. Hall[1], Paula Parra-Bueno[2], Teresa A. Milner[3,4], Enhui Pan[1], Tal Laviv[2], Barbara L. Hempstead[3,5], Ryohei Yasuda[1,2] & James O. McNamara[1]

Brain-derived neurotrophic factor (BDNF) and its receptor TrkB are crucial for many forms of neuronal plasticity[1–6], including structural long-term potentiation (sLTP)[7,8], which is a correlate of an animal's learning[7,9–12]. However, it is unknown whether BDNF release and TrkB activation occur during sLTP, and if so, when and where. Here, using a fluorescence resonance energy transfer-based sensor for TrkB and two-photon fluorescence lifetime imaging microscopy[13–16], we monitor TrkB activity in single dendritic spines of CA1 pyramidal neurons in cultured murine hippocampal slices. In response to sLTP induction[9,14–16], we find fast (onset < 1 min) and sustained (>20 min) activation of TrkB in the stimulated spine that depends on NMDAR (*N*-methyl-D-aspartate receptor) and CaMKII signalling and on postsynaptically synthesized BDNF. We confirm the presence of postsynaptic BDNF using electron microscopy to localize endogenous BDNF to dendrites and spines of hippocampal CA1 pyramidal neurons. Consistent with these findings, we also show rapid, glutamate-uncaging-evoked, time-locked BDNF release from single dendritic spines using BDNF fused to superecliptic pHluorin[17–19]. We demonstrate that this postsynaptic BDNF–TrkB signalling pathway is necessary for both structural and functional LTP[20]. Together, these findings reveal a spine-autonomous, autocrine signalling mechanism involving NMDAR–CaMKII-dependent BDNF release from stimulated dendritic spines and subsequent TrkB activation on these same spines that is crucial for structural and functional plasticity.

To address the role of BDNF–TrkB signalling in sLTP, we developed a fluorescence resonance energy transfer (FRET)-based sensor for TrkB consisting of two components: (1) TrkB fused to monomeric enhanced green fluorescent protein (TrkB–eGFP), and (2) an SH2 domain of the TrkB binding partner phospholipase Cγ1 (PLC-γ1)[21] fused to two copies of monomeric red fluorescent protein-1 (mRFP1–PLC–mRFP1; Fig. 1a and Supplementary Information). After TrkB activation via phosphorylation of Tyr816, the affinity of mRFP1–PLC–mRFP1 for TrkB–eGFP increases[21], thereby allowing FRET to occur between the fluorophores (Supplementary Information). We validated the sensor in HeLa cells and cultured cortical neurons by showing it to be sensitive to BDNF, specific for Tyr816phosphorylation, and reversible when imaged by two-photon fluorescence lifetime imaging microscopy (2pFLIM) (Extended Data Fig. 1, Supplementary Information). Furthermore, we demonstrated that the sensor could functionally replace endogenous TrkB in neurons of cultured hippocampal slices (Extended Data Fig. 2, Supplementary Information).

Using this sensor, we biolistically transfected cultured rat hippocampal slices and imaged CA1 pyramidal neurons with 2pFLIM. In response to glutamate uncaging targeted to a single dendritic spine (30 pulses at 0.5 Hz), spine volume rapidly increased by ~220% (transient phase) before relaxing to an increased state of ~90% lasting at least 60 min (sustained phase; Fig. 1b, c, Extended Data Fig. 3a)—changes independent of protein synthesis (Extended Data Fig. 3c, d) and largely consistent with previous descriptions of sLTP (Extended Data Fig. 3a, Supplementary Information)[9,14–16]. At the same time, TrkB rapidly activated in the stimulated spine, peaking at ~1–2 min and remaining elevated for at least 60 min (Fig. 1b, d, e, Extended Data Figs 3b, 4 and Supplementary Information). For the first 30–60 s after the onset of glutamate uncaging, this activation was largely restricted to the stimulated spine (Fig. 1d–f, Supplementary Information). However, with time, TrkB activation in adjacent regions slowly increased, suggesting spreading of TrkB activation (Fig. 1d–f). The validity of the observed signal was confirmed by its dependence on kinase activity and Tyr816 phosphorylation and independence of sensor concentration and temperature (Extended Data Figs 5–7, Supplementary Information).

To explore mechanisms underlying this TrkB activation, we asked whether it required NMDAR-mediated Ca^{2+} influx[7,9,15,16] and subsequent CaMKII activation[9,15]. Application of either the NMDAR inhibitor D-2-amino-5-phosphonovalerate (D-AP5; 100 μM) or the CaMKII inhibitor CN21 (ref. 22; 10 μM) impaired TrkB activation during the transient and sustained phases of sLTP while also inhibiting spine volume change (Fig. 2a–d), suggesting that TrkB is in part downstream of both NMDAR and CaMKII activation.

Next, we asked whether BDNF contributes to this TrkB activation. Using the extracellular BDNF scavenger TrkB-Ig (6–8 μg ml^{-1}), we found impaired TrkB activation throughout sLTP with a similar impairment of spine volume change (Fig. 2e–h), suggesting a crucial role for BDNF in mediating glutamate-uncaging-induced TrkB activation. To examine the cellular source of BDNF underlying this activation, we sparsely transfected the sensor with Cre recombinase in slices from *Bdnf*$^{fl/fl}$ mice, thereby selectively knocking-out BDNF synthesized in the postsynaptic cell, a perturbation without detectable effect on basal spine morphology[23,40,41] (Extended Data Fig. 8a, b). This manipulation attenuated glutamate-uncaging-evoked TrkB activation and sLTP (Fig. 2i–l) while leaving CaMKII activation intact (Extended Data Fig. 8c–f). These results implicate autocrine BDNF as one mechanism underlying TrkB activation during sLTP: additional mechanisms could include other sources of BDNF (pre-synaptic, paracrine) or non-neurotrophin TrkB activators (such as zinc)[24].

The dependence of glutamate-uncaging-induced TrkB activation on postsynaptically synthesized BDNF controversially suggests the existence of BDNF in dendrites or spines[25]. To provide more direct evidence, we used electron microscopy to examine BDNF localization in a previously characterized mouse line in which a C-terminal haemagglutinin (HA) epitope tag was added to the *Bdnf* coding sequence (*Bdnf-HA*)[26]. Using highly sensitive antibodies against the HA-tag, we found BDNF not only in axons but also in dendrites and spines of CA1 pyramidal cells of these mice (Fig. 3).

[1]Neurobiology Department, Duke University Medical Center, Research Drive, Durham, North Carolina 27710, USA. [2]Max Planck Florida Institute for Neuroscience, 1 Max Planck Way, Florida 33458, USA. [3]Feil Family Brain and Mind Research Institute, Weill Cornell Medicine, 407 East 61st St, New York 10065, USA. [4]Laboratory of Neuroendocrinology, The Rockefeller University, 1230 York Avenue, New York 10065, USA. [5]Department of Medicine, Weill Cornell Medicine, 1300 York Avenue, New York 10065, USA. †Present address: Neurobiology Section, Center for Neural Circuits and Behavior, and Department of Neurosciences, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093, USA.
*These authors contributed equally to this work.

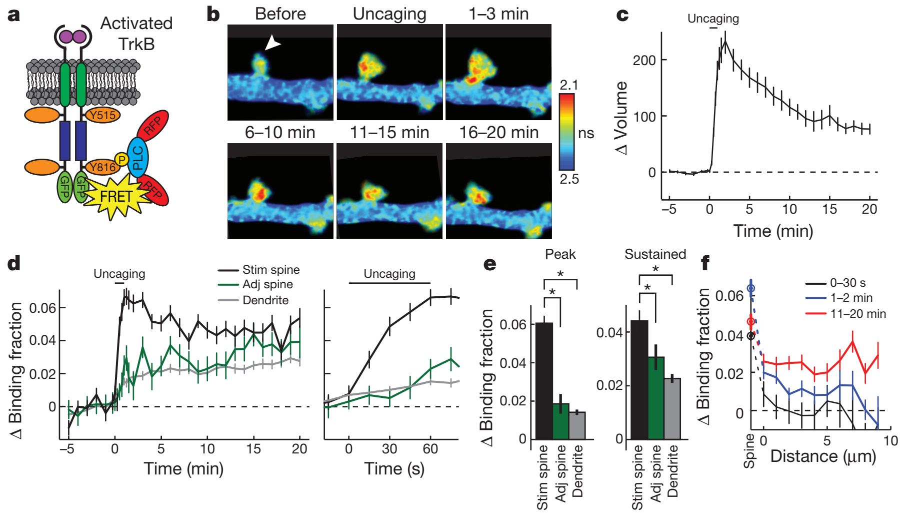


Figure 1 | sLTP induces rapid, persistent, and largely spine specific TrkB activation. a, Sensor design. **b**, 2pFLIM images of TrkB activation averaged across indicated time points. Arrowhead represents point of uncaging. Warmer colours indicate shorter lifetimes and higher TrkB activity. Image size is 6.8 × 4.4 μm. **c**, Time course of volume change for the stimulated spine. $n = 50$ cells/54 spines. **d**, **e**, Time course (**d**) and quantification (**e**) of peak (1.25–2 min) and sustained (10–20 min) activation for experiments in **c** measured as the change in sensor binding fraction in stimulated spines, adjacent spines and dendrites. Right panel in **d** shows magnified time course. $n = 50/54$ for stimulated spines and dendritic shafts, and 50/59 for adjacent spines (cells/spines). **f**, Spatial profile of TrkB activation—change in binding fraction of the dendrite plotted as a function of the distance from the stimulated spine. $n = 48/52$ (cells/stimulated spines). Data are mean ± s.e.m. *$P < 0.05$, analysis of variance (ANOVA) with Dunnet's test.

The localization of BDNF to dendritic spines, together with the rapid kinetics and initial spine restriction of glutamate-uncaging-induced TrkB activation, suggested an equally rapid release of BDNF from dendritic spines during the transient phase of sLTP. To assess this possibility, we used biolistics to transfect CA1 pyramidal cells with full-length BDNF containing the pH-sensitive fluorophore superecliptic pHluorin (SEP) fused to its C terminus, thus allowing visualization of exocytosed BDNF[18,19] (Extended Data Fig. 9a,

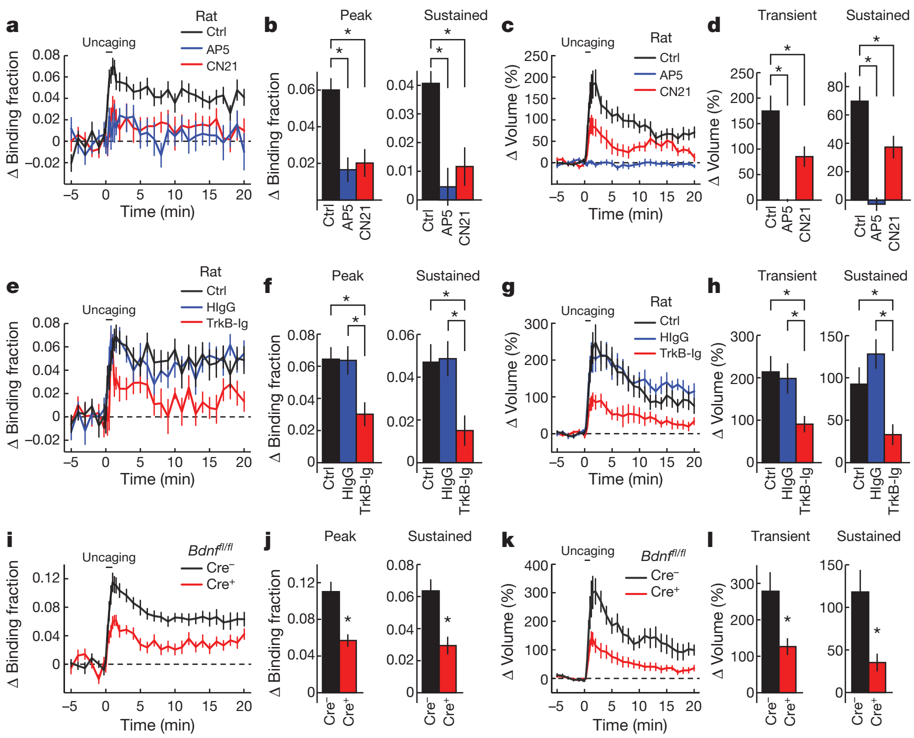


Figure 2 | TrkB activation during sLTP depends on NMDAR–CaMKII signalling and postsynaptic BDNF. a, **b**, Time course (**a**) and quantification (**b**) of peak and sustained TrkB activation in stimulated spines in the presence of pharmacological inhibitors. Ctrl, control; AP5 denotes an NMDAR inhibitor; CN21 denotes a CaMKII inhibitor. $n = 19/19$ control, 6/10 AP5, and 7/16 CN21 (cells/spines). **c**, **d**, Time course (**c**) and quantification (**d**) of transient (1–2 min) and sustained (10–20 min) spine volume change for experiments in **a** and **b**. **e**–**h**, Similar experiments to **a**–**d** but with different pharmacological conditions. HIgG, human IgG; TrkB-Ig, an extracellular scavenger of BDNF. $n = 16/18$ control, 6/11 HIgG, and 8/14 TrkB-Ig (cells/spines). **i**–**l**, Similar experiments to **a**–**d** but in *Bdnf*$^{fl/fl}$ hippocampal slices transfected with the TrkB sensor with or without Cre-recombinase (Cre^- or Cre^+, respectively). $n = 7/15$ Cre^- and 9/17 Cre^+ (cells/spines). Data are mean ± s.e.m. *$P < 0.05$, ANOVA with Dunnet's test (**b**, **d**), Tukey's test (**f**, **h**) or a two-tailed *t*-test (**j**, **l**).

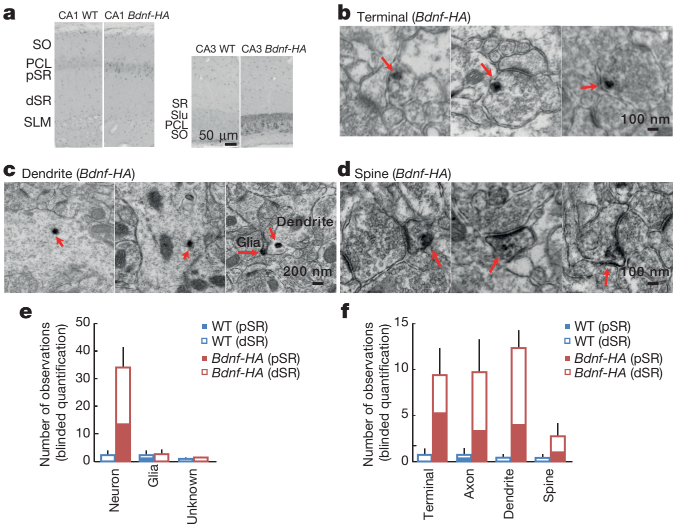


Figure 3 | Endogenous BDNF localizes to axons, dendrites and dendritic spines. **a**, Immunoperoxidase labelling of HA in hippocampal areas CA1 (left) and CA3 (right) from *Bdnf-HA* and wild-type (WT) mice, visualized by light microscopy. dSR, distal stratum radiatum; PCL, principal cell layer; pSR, proximal stratum radiatum; SLM, stratum lacunosum-moleculare; SO, stratum oriens. **b**–**d**, Immunoperoxidase labelling of HA in CA1 pyramidal neuron axon terminals (**b**), dendrites (**c**), and dendritic spines (**d**) of *Bdnf-HA* mice, visualized by electron microscopy. **e**, **f**, Quantification of observed immunoperoxidase labelling of HA in various cellular types (**e**) and subcellular compartments (**f**) in proximal and distal stratum radiatum in hippocampal slices from wild-type and *Bdnf-HA* mice. $n = 3$ animals each.

Supplementary Information). In response to glutamate uncaging, we observed an increase in SEP fluorescence largely restricted to the stimulated spine (Fig. 4a, b, Supplementary Video 1), with two distinct kinetic profiles. The first was a transient, spike-like increase time-locked to each uncaging pulse (Fig. 4a–c), perhaps correlating with activity-induced BDNF release. The second was a slow increase in fluorescence commencing with the start of uncaging and peaking at its end, perhaps due to extracellular accumulation of released BDNF–SEP (Fig. 4a, b). After termination of uncaging, SEP fluorescence decayed back to baseline over the course of 10 min (Extended Data Fig. 9e).

Several lines of evidence suggested that these transient, spike-like increases in SEP fluorescence were in fact due to BDNF release from the spine. First, the observed fluorescence signal depended on pH, exocytosis, and BDNF sorting machinery[27] (Fig. 4c, d, Extended Data Fig. 9d, Supplementary Information). Second, expression of BDNF–SEP rescued structural plasticity in the setting of postsynaptic BDNF knockout, thus suggesting that it could functionally replace endogenous BDNF in neurons (Fig. 4e, f, Supplementary Information). Third, the observed BDNF–SEP signal was independent of the presence of endogenous postsynaptic BDNF (Fig. 4g, h). Fourth, the kinetic profile of the signal paralleled the time course of TrkB activation, as one would expect for BDNF release (Extended Data Fig. 4). Collectively, these results suggest that the observed increase in SEP signal probably reports glutamate-dependent exocytosis and release of BDNF from stimulated spines.

To explore mechanisms underlying glutamate-induced BDNF release, we inhibited NMDARs (with AP5) or AMPARs (with NBQX) individually and together as well as inhibiting CaMKII with CN21 (Fig. 4c, d). We found the SEP signal to be largely blocked (AP5), unaffected (NBQX), completely blocked (AP5 plus NBQX), and partially blocked (CN21) by these perturbations (Fig. 4c, d). These findings suggest that BDNF release from spines is largely NMDAR–CaMKII dependent, consistent with our results for TrkB activation.

The converging evidence implicating autocrine BDNF–TrkB signalling in sLTP led us to ask whether this pathway was also involved

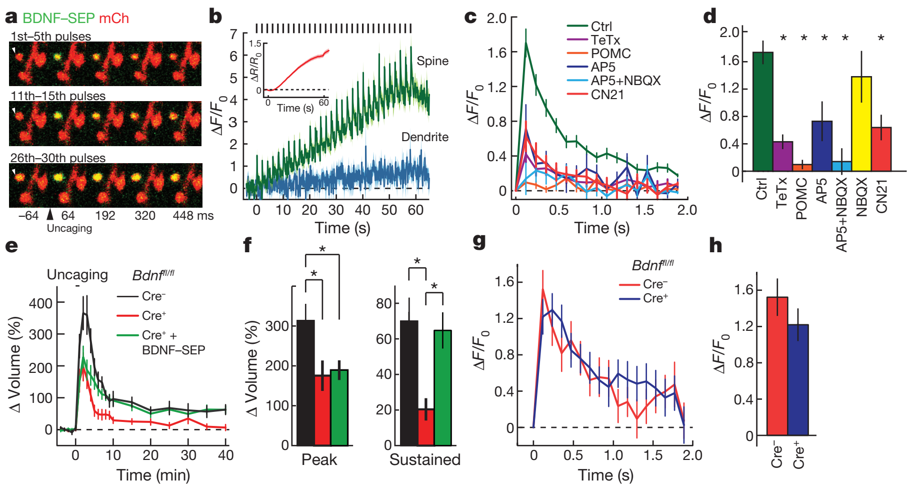


Figure 4 | Glutamate uncaging induces rapid release of postsynaptic BDNF. **a**, Two-photon images of glutamate-uncaging-evoked changes in BDNF–SEP fluorescence in dendritic spines of CA1 hippocampal neurons. Each row represents the uncaging-triggered average of the BDNF–SEP signal in response to individual uncaging pulses for the designated time window. Image size is 3.9×5.5 µm. **b**, Averaged time course of BDNF–SEP fluorescence change in spines and adjacent dendritic shafts in response to glutamate uncaging (timing of glutamate pulses indicated by black bars (top)). Inset shows the change in mCherry (mCh) fluorescence (red) in response to glutamate uncaging, indicative of spine volume change (sLTP). $n = 26/187$ (cells/spines). **c**, Uncaging-triggered average of the increase in BDNF–SEP fluorescence with glutamate uncaging. TeTx, neurons transfected with tetanus toxin, an inhibitor of exocytosis; POMC, neurons transfected with the POMC peptide, an inhibitor of activity-dependent BDNF release. $n = 31/218$ control, 6/82 TeTx, 2/29 POMC, 3/50 AP5, 2/46 AP5 + NBQX, 4/40 NBQX, and 7/88 CN21 (cells/spines). **d**, Peak of the uncaging-triggered averaged increase of BDNF–SEP fluorescence in **c**. **e**, Time course of glutamate-uncaging-induced spine volume change for *Bdnf*$^{fl/fl}$ hippocampal slices transfected with eGFP (Cre$^-$), eGFP plus Cre (Cre$^+$), or eGFP, Cre and BDNF–SEP. $n = 9/13$ Cre$^-$, 6/11 Cre$^+$ and 8/13 Cre$^+$ plus BDNF–SEP (cells/spines). **f**, Transient (1–2 min) and sustained (10–40 min) spine volume change for experiments in **e**. **g**, **h**, Similar experiments to **c** and **d** but in *Bdnf*$^{fl/fl}$ hippocampal slices in the absence or presence of Cre. $n = 10/105$ Cre$^-$ and 15/132 Cre$^+$ (cells/spines). Data are mean ± s.e.m. See Extended Data Fig. 9h for data in **d** represented as median ± interquartile interval. *$P < 0.05$, Kruskal–Wallis test with Dunn's test (**d**) or an ANOVA with Tukey's test (**f**).

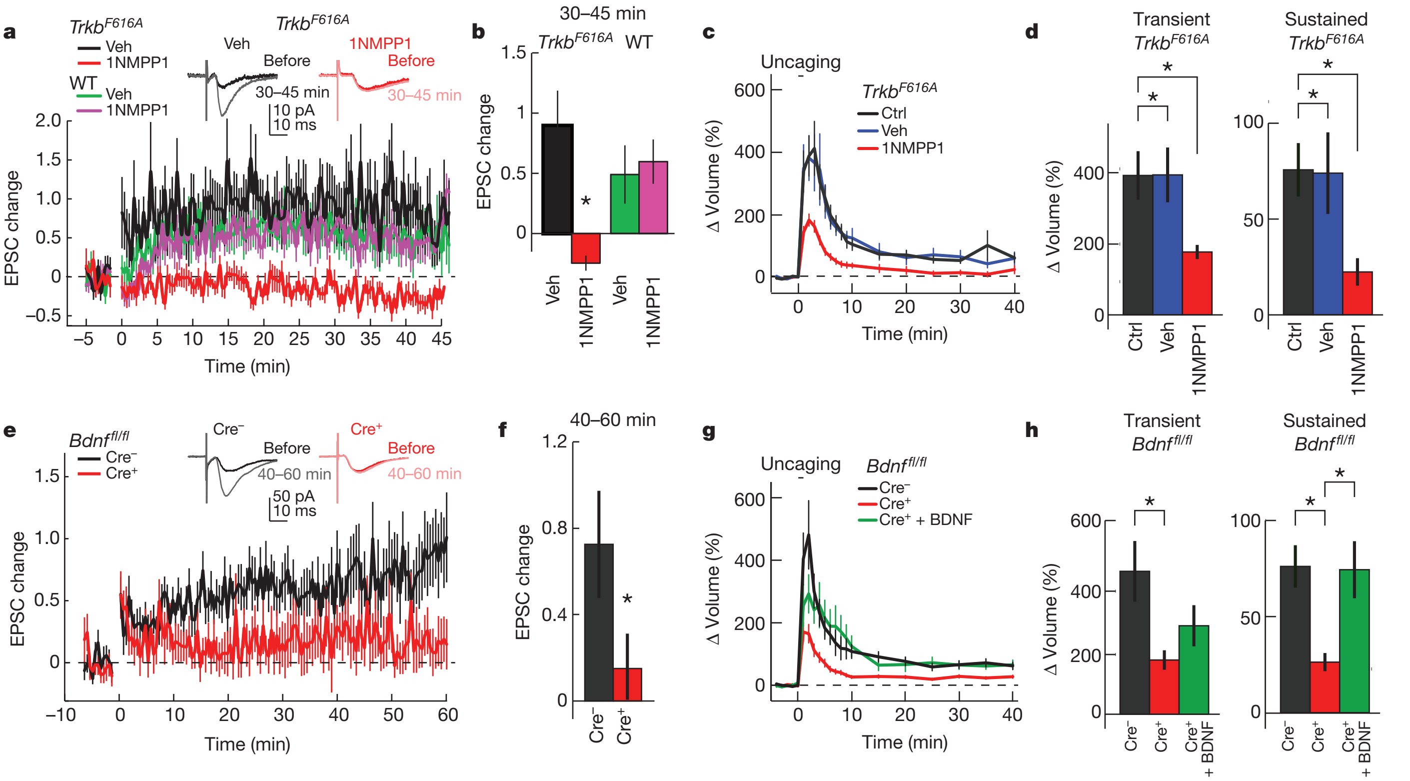


Figure 5 | Functional and structural LTP depends on postsynaptic BDNF–TrkB signalling. a, b, Time course (**a**) and quantification (**b**; 30–45 min) of excitatory postsynaptic current (EPSC) change recorded in CA1 pyramidal cells of hippocampal slices from *Trkb*^{F616A} and wild-type mice, before and after LTP induction in the presence of vehicle or 1NMPP1. Representative traces of *Trkb*^{F616A} slices with vehicle or 1NMPP1 are shown above the graphs. $n = 11$ *Trkb*^{F616A} vehicle, 10 *Trkb*^{F616A} 1NMPP1, 11 wild-type vehicle, and 13 wild-type 1NMPP1 (cells). **c, d,** Time course (**c**) and quantification (**d**) of transient and sustained glutamate-uncaging-induced spine volume change for *Trkb*^{F616A} hippocampal slices in the absence or presence of vehicle or 1NMPP1. $n = 20/20$ control, 10/13 vehicle, and 16/20 1NMPP1 (cells/spines). **e–h,** Similar experiments to **a** and **b** but from *Bdnf*$^{fl/fl}$ mice infected with or without Cre. Representative traces are shown above the graphs. $n = 20$ Cre$^-$ and 19 Cre$^+$ (cells). **g, h,** Time course (**g**) and quantification (**h**) of transient and sustained glutamate-uncaging-induced spine volume change for *Bdnf*$^{fl/fl}$ hippocampal slices transfected with eGFP or eGFP plus Cre. For Cre$^+$ plus BDNF, Cre-positive cells were treated with BDNF for 10 min before glutamate uncaging. $n = 13/14$ Cre$^-$, 22/32 Cre$^+$, and 6/7 Cre$^+$ plus BDNF (cells/spines). Data are mean ± s.e.m. $*P < 0.05$, two-tailed *t*-test (**b, f**) or ANOVA with Tukey's test (**d, h**).

in functional LTP (fLTP) at the CA3–CA1 synapse. To address this question, we induced fLTP by pairing low-frequency Schaffer collateral axon stimulation with depolarization of single CA1 pyramidal cells through whole-cell patch clamping. First, we examined the role of TrkB by using knock-in mice containing a point mutation in the TrkB kinase domain (F616A; *Trkb*^{F616A}, also known as *Ntrk2*^{F616A}), rendering the mutant TrkB uniquely susceptible to inhibition by the small molecule 1NMPP1 (ref. 28). 1NMPP1 inhibited both sLTP (in cultured slices) and fLTP (in acute slices) in slices isolated from *Trkb*^{F616A} but not wild-type mice, revealing a requirement for TrkB kinase (Fig. 5a–d). In addition, scavenging extracellular BDNF with TrkB-Ig ($2\,\mu g\,ml^{-1}$) impaired both sLTP (in cultured slices) and fLTP (in acute slices) (Extended Data Fig. 10a–d), implicating BDNF as one mechanism underlying TrkB activation in this context.

To determine whether autocrine BDNF–TrkB signalling in particular contributed to these forms of plasticity, we knocked out BDNF in a small population of CA1 pyramidal cells using either *in utero* infection of adeno-associated virus encoding *synapsin-Cre* in *Bdnf*$^{fl/fl}$ mice for fLTP in acute slices or biolistic transfection of Cre in organotypic slices prepared from *Bdnf*$^{fl/fl}$ mice for sLTP. The knockout of postsynaptic BDNF impaired both fLTP and sLTP, the latter of which was rescued by bath-applied BDNF ($20\,ng\,ml^{-1}$ for 10 min) (Fig. 5e–h). Collectively, these results reveal a requirement of a cell-autonomous, postsynaptic BDNF release and subsequent activation of postsynaptic TrkB for both structural and functional synaptic plasticity (Extended Data Fig. 10e, Supplementary Information).

Overall, we have described an autocrine signalling system within a single spine achieved by rapid BDNF release from the stimulated spine and subsequent TrkB activation on the same spine that, potentially in cooperation with other sources of BDNF and activators of TrkB, is essential for both structural and functional plasticity.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 5 April; accepted 15 August 2016.
Published online 28 September 2016.

1. Lohof, A. M., Ip, N. Y. & Poo, M. M. Potentiation of developing neuromuscular synapses by the neurotrophins NT-3 and BDNF. *Nature* **363,** 350–353 (1993).
2. Kang, H., Welcher, A. A., Shelton, D. & Schuman, E. M. Neurotrophins and time: different roles for TrkB signaling in hippocampal long-term potentiation. *Neuron* **19,** 653–664 (1997).
3. Minichiello, L. *et al.* Essential role for TrkB receptors in hippocampus-mediated learning. *Neuron* **24,** 401–414 (1999).
4. Figurov, A., Pozzo-Miller, L. D., Olafsson, P., Wang, T. & Lu, B. Regulation of synaptic responses to high-frequency stimulation and LTP by neurotrophins in the hippocampus. *Nature* **381,** 706–709 (1996).
5. Korte, M. *et al.* Hippocampal long-term potentiation is impaired in mice lacking brain-derived neurotrophic factor. *Proc. Natl Acad. Sci. USA* **92,** 8856–8860 (1995).
6. Kovalchuk, Y., Hanse, E., Kafitz, K. W. & Konnerth, A. Postsynaptic induction of BDNF-mediated long-term potentiation. *Science* **295,** 1729–1734 (2002).
7. Tanaka, J. *et al.* Protein synthesis and neurotrophin-dependent structural plasticity of single dendritic spines. *Science* **319,** 1683–1687 (2008).

8. Lai, K.-O. O. *et al.* TrkB phosphorylation by Cdk5 is required for activity-dependent structural plasticity and spatial memory. *Nat. Neurosci.* **15,** 1506–1515 (2012).
9. Matsuzaki, M., Honkura, N., Ellis-Davies, G. C. & Kasai, H. Structural basis of long-term potentiation in single dendritic spines. *Nature* **429,** 761–766 (2004).
10. Okamoto, K., Nagai, T., Miyawaki, A. & Hayashi, Y. Rapid and persistent modulation of actin dynamics regulates postsynaptic reorganization underlying bidirectional plasticity. *Nat. Neurosci.* **7,** 1104–1112 (2004).
11. Kim, I. H. *et al.* Disruption of Arp2/3 results in asymmetric structural plasticity of dendritic spines and progressive synaptic and behavioral abnormalities. *J. Neurosci.* **33,** 6081–6092 (2013).
12. Kim, I. H., Wang, H., Soderling, S. H. & Yasuda, R. Loss of Cdc42 leads to defects in synaptic plasticity and remote memory recall. *eLife* **3,** e02839 (2014).
13. Yasuda, R. Imaging spatiotemporal dynamics of neuronal signaling using fluorescence resonance energy transfer and fluorescence lifetime imaging microscopy. *Curr. Opin. Neurobiol.* **16,** 551–561 (2006).
14. Harvey, C. D., Yasuda, R., Zhong, H. & Svoboda, K. The spread of Ras activity triggered by activation of a single dendritic spine. *Science* **321,** 136–140 (2008).
15. Lee, S.-J. R. J., Escobedo-Lozoya, Y., Szatmari, E. M. & Yasuda, R. Activation of CaMKII in single dendritic spines during long-term potentiation. *Nature* **458,** 299–304 (2009).
16. Murakoshi, H., Wang, H. & Yasuda, R. Local, persistent activation of Rho GTPases during plasticity of single dendritic spines. *Nature* **472,** 100–104 (2011).
17. Miesenböck, G., De Angelis, D. A. & Rothman, J. E. Visualizing secretion and synaptic transmission with pH-sensitive green fluorescent proteins. *Nature* **394,** 192–195 (1998).
18. Matsuda, N. *et al.* Differential activity-dependent secretion of brain-derived neurotrophic factor from axon and dendrite. *J. Neurosci.* **29,** 14185–14198 (2009).
19. Dean, C. *et al.* Synaptotagmin-IV modulates synaptic function and long-term potentiation by regulating BDNF release. *Nat. Neurosci.* **12,** 767–776 (2009).
20. Hedrick, N. G. *et al.* Rho GTPase complementation underlies BDNF-dependent homo- and heterosynaptic plasticity. *Nature* http://dx.doi.org/10.1038/nature19784 (2016).
21. Middlemas, D. S., Meisenhelder, J. & Hunter, T. Identification of TrkB autophosphorylation sites and evidence that phospholipase C-γ1 is a substrate of the TrkB receptor. *J. Biol. Chem.* **269,** 5458–5466 (1994).
22. Vest, R. S., Davies, K. D., O'Leary, H., Port, J. D. & Bayer, K. U. Dual mechanism of a natural CaMKII inhibitor. *Mol. Biol. Cell* **18,** 5024–5033 (2007).
23. Lu, W. *et al.* Subunit composition of synaptic AMPA receptors revealed by a single-cell genetic approach. *Neuron* **62,** 254–268 (2009).
24. Huang, Y. Z., Pan, E., Xiong, Z.-Q. Q. & McNamara, J. O. Zinc-mediated transactivation of TrkB potentiates the hippocampal mossy fiber-CA3 pyramid synapse. *Neuron* **57,** 546–558 (2008).
25. Dieni, S. *et al.* BDNF and its pro-peptide are stored in presynaptic dense core vesicles in brain neurons. *J. Cell Biol.* **196,** 775–788 (2012).
26. Yang, J. *et al.* Neuronal release of proBDNF. *Nat. Neurosci.* **12,** 113–115 (2009).
27. Lou, H. *et al.* Sorting and activity-dependent secretion of BDNF require interaction of a specific motif with the sorting receptor carboxypeptidase E. *Neuron* **45,** 245–255 (2005).
28. Chen, X. *et al.* A chemical-genetic approach to studying neurotrophin signaling. *Neuron* **46,** 13–21 (2005).
40. He, X.-P. *et al.* Conditional deletion of TrkB but not BDNF prevents epileptogenesis in the kindling model. *Neuron* **43,** 31–42 (2004).
41. Luikart, B. W. *et al.* TrkB has a cell-autonomous role in the establishment of hippocampal Schaffer collateral synapses. *J Neurosci.* **25,** 3774–3786 (2005).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank A. West and Y. Huang for critical discussion. This work was supported by grants from the National Institutes of Health (F31NS078847 (S.C.H.), R01NS068410 (R.Y.), DP1NS096787 (R.Y.), R01NS05621 (J.O.M.), R01MH080047 (R.Y.), R01DA08259 (T.A.M.), R01HL098351 (T.A.M.), P01HL096571 (T.A.M.), and R01NS030687 (B.L.H.)), the Wakeman Fellowship (S.C.H.), and Human Frontier Science Program (T.L.).

Author Contributions S.C.H., N.G.H., R.Y. and J.O.M. designed experiments; S.C.H. and N.G.H. collected and analysed imaging data with assistance from C.E.H.; P.P.-B. and E.P. collected and analysed patch clamp data; T.A.M. collected electron microscopic images and analysed them with B.L.H.; T.L. performed in utero viral injections; S.C.H., N.G.H., R.Y. and J.O.M. analysed remaining data and wrote the paper. All authors discussed results and comments on this manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.Y (Ryohei.Yasuda@mpfi.org).

Reviewer Information *Nature* thanks B. Bingol, H. Zhang and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Reagents. Human recombinant BDNF and human recombinant β-NGF were purchased from Millipore, K252a and D-2-amino-5-phosphonovalerate (D-AP5) and 2,3-dihydroxy-6-nitro-7-sulfamoyl-benzo[f]quinoxaline-2,3-dione (NBQX) were from Tocris, human-IgG was from Sigma, and 1'-naphthylmethyl-4-amino-1-*tert*-butyl-3-(*p*-methylphenyl)pyrazolo[3,4-*d*] pyrimidine (1NMPP1) was from Santa Cruz and Shanghai Institute of Materia Medica, Chinese Academy of Sciences. TrkB-Ig was a gift from Regeneron and the tat-CN21 peptide (YGRKKRRQRRRKRPPKLGQIGRSKRVVIEDDR) was synthesized by GenScript.

Plasmids. TrkB–eGFP was prepared by inserting the coding sequence of mouse TrkB (obtained from a previously described plasmid[29]) into pEGFP-N1 (Clontech) containing the A206K monomeric mutation in eGFP and the CAG promoter[30]. The linker between TrkB and eGFP is TGRH. mRFP1–PLC–mRFP1 was prepared by inserting the coding sequence for the C-terminal SH2 domain of human PLC-γ1 (659–769; obtained from full-length, human PLC-γ1 purchased from Origene) into a tandem-mRFP1 plasmid containing the CAG promoter. The linkers between the mRFP1s and PLC-γ1 (659–769) are RSRAQASNS for the N terminus and GSG for the C terminus. $\text{TrkB}^{\text{Y816F}}$–eGFP was prepared by introducing a point mutation using the Site-Directed Mutagenesis Kit (Stratagene). Tandem mCherry (mCh–mCh) was generated as previously described[16]. HA–BDNF–Flag was a gift from A. West. The coding sequence for SEP (obtained from SEP-GluA1; ref. 31) was incorporated onto the 3′ end of HA–BDNF–Flag to generate HA-BDNF–Flag-SEP. HA–BDNF–Flag–mRFP1 was generated in a similar fashion. A plasmid containing mCh-IRES-TeTX was a gift from M. Ehlers. POMC-mCh was generated by amplifying the POMC peptide (MWCLESSQCQDLTTESNLLACIRACRLDL)[27] using overhang PCR with a C-terminal linker (GGGGGGGGGGGGGGG GGGGGGGGGMADQLTEEWHRGTAGPGS). This amplicon was then inserted into the tandem mCh plasmid by replacing the coding sequence of the first mCh.

Animals. All animal procedures were approved by the Duke University School of Medicine Animal Care and Use Committee, Max Planck Florida Institute for Neuroscience, and Weill Cornell Medical College Institutional Animal Care and Use Committees and were conducted in accordance with the NIH Guide for the Care and Use of Laboratory Animals.

We used both male and female rats and mice. Rats and C57/B6 mice were obtained from Charles River, *Trkb*^{F616A} mutant mice were provided by D. Ginty[28], *Bdnf*$^{fl/fl}$ and *Trkb*$^{fl/fl}$ mice were provided by L. Parada[32], and *Bdnf-HA* mice were generated as previously described[26]. The genotype of each animal used was verified before and after preparing slices using PCR of genomic DNA isolated from tail DNA before and slice samples after.

Preparation of HeLa cells. HeLa cells were obtained from the Duke University Cell Culture Facility. These cells had been authenticated using short-tandem repeat profiling and evaluated for mycoplasma contamination. Cells were cultured and maintained as previously described[16]. Cells were transfected with Lipofectamine 2000 using the manufacturer's protocol (Invitrogen). Concentrations used were 0.5 μl ml^{-1} Lipofectamine and 1 μg ml^{-1} total cDNA (1:1 ratio of TrkB–eGFP to mRFP1–PLC–mRFP1 DNA). Then, 24–48 h later, culture media was replaced with HEPES-buffered ACSF for imaging (HACSF; 20 mM HEPES, 130 mM NaCl, 2 mM $NaHCO_3$, 25 mM D-glucose, 2.5 mM KCl and 1.25 mM NaH_2PO_4; adjusted to pH 7.4 and 310 mOsm). After a 30-min equilibration period, transfected cells were imaged using 2pFLIM as described below. Cell stimulation was performed by directly adding BDNF or vehicle to the HACSF bathing the cells.

Preparation of mixed cortical cultures. Mixed cortical cultures were prepared as described previously[33] and transfected with Lipofectamine 2000 using a modified protocol. For transfection of neurons in 3.5 cm dishes, 1 μl Lipofectamine was mixed with 1 μg of plasmid DNA (1 μg per construct transfected) in 100 μl of culture media for 20 min. Culture media was removed from the 3.5 cm dish until only 1 ml remained. The Lipofectamine/DNA solution was added to the neurons for 45 min. At this point, all the media was removed and replaced with 2 ml conditioned culture media. After 24–48 h, culture media was replaced with HACSF. To stimulate cells, we added BDNF or NGF directly to the HACSF bathing the cells. 30 min after stimulation, we added K252a to the HACSF.

Preparation of organotypic hippocampal slices. Cultured hippocampal slices were prepared from post-natal day 5–7 rats or mice, as previously described[34], in accordance with the animal care and use guidelines of Duke University Medical Center. After 5–12 days in culture, CA1 pyramidal neurons were transfected with biolistic gene transfer using gold beads (12 mg; Biorad) coated with plasmids containing 20–40 μg of total cDNA (TrkB sensor: 15 μg TrkB–eGFP and 15 μg mRFP1–PLC–mRFP1; TrkB sensor plus mCh: 5 μg TrkB–eGFP, 5 μg mRFP1–PLC–mRFP1, and 20 μg mCh–mCh; TrkB sensor plus mCh and Cre: 5 μg TrkB–GFP, 5 μg mRFP1–PLC–mRFP1, 5 μg tdTom-Cre, and 15 μg mCh–mCh; BDNF–SEP plus mCh: 20 μg BDNF–SEP and 10 μg mCh–mCh; BDNF–SEP plus TeTX: 20 μg BDNF–SEP and 10 μg mCh-IRES-TeTX; BDNF–SEP plus POMC: 20 μg BDNF–SEP and 10 μg POMC–mCh; eGFP: 20 μg eGFP; and eGFP plus Cre: 10 μg eGFP plus 10 μg tdTom-Cre). Neurons expressing the TrkB sensor were imaged 12–48 h after transfection. Neurons expressing the TrkB sensor with mCh or mCh plus Cre were imaged 5–7 days after transfection. The addition of the mCh proved critical in limiting the TrkB sensor expression thereby allowing neurons to survive longer with the sensor present. Neurons expressing only eGFP were imaged 1–7 days after transfection. Neurons expressing eGFP plus Cre were imaged 5–9 days after transfection.

2pFLIM. FRET imaging using a custom-built two-photon fluorescence lifetime imaging microscope was performed as previously described[13,35]. Two-photon imaging was performed using a Ti-sapphire laser (MaiTai, Spectraphysics) tuned to a wavelength of 920 nm, allowing simultaneous excitation of eGFP, mRFP1 and mCh. All samples were imaged using <2 mW laser power measured at the objective. Fluorescence emission was collected using an immersion objective (60×, numerical aperture 0.9, Olympus), divided with a dichroic mirror (565 nm), and detected with two separate photoelectron multiplier tubes (PMTs) placed downstream of two wavelength filters (Chroma, HQ510-2p to select for green and HQ620/90-2p to select for red). The green channel was fitted with a PMT having a low transfer time spread (H7422-40p; Hamamatsu) to allow for fluorescence lifetime imaging, while the red channel was fitted with a wide-aperture PMT (R3896; Hamamatsu). Photon counting for fluorescence lifetime imaging was performed using a time-correlated single photon counting board (SPC-150; Becker and Hickl) controlled with custom software[13], while the red channel signal was acquired using a separate data acquisition board (PCI-6110) controlled with Scanimage software[36].

Two-photon glutamate uncaging. A second Ti-sapphire laser tuned at a wavelength of 720 nm was used to uncage 4-methoxy-7-nitroindolinyl-caged-L-glutamate (MNI-caged glutamate) in extracellular solution with a train of 4–6 ms, 4–5 mW pulses (30 times at 0.5 Hz) near a spine of interest. Experiments were performed in Mg^{2+} free artificial cerebral spinal fluid (ACSF; 127 mM NaCl, 2.5 mM KCl, 4 mM $CaCl_2$, 25 mM $NaHCO_3$, 1.25 mM NaH_2PO_4 and 25 mM glucose) containing 1 μM tetrodotoxin (TTX) and 4 mM MNI-caged L-glutamate aerated with 95% O_2 and 5% CO_2. Experiments were performed at 24–26 °C (room temperature) or 30–32 °C using a heating block holding the ACSF container. Temperature measurements were made from ACSF within the perfusion chamber holding the slice.

2pFLIM data analyses. To measure the fraction of donor bound to acceptor, we fit a fluorescence lifetime curve summing all pixels over a whole image with a double exponential function convolved with the Gaussian pulse response function:

$$F(t) = F_0[P_D H(t, t_0, \tau_D, \tau_G) + P_{AD} H(t, t_0, \tau_{AD}, \tau_G)]$$

in which τ_{AD} is the fluorescence lifetime of donor bound with acceptor, P_D and P_{AD} are the fraction of free donor and donor bound with acceptor, respectively, and $H(t)$ is a fluorescence lifetime curve with a single exponential function convolved with the Gaussian pulse response function:

$$H(t, t_0, \tau_D, \tau_G) = \frac{1}{2}\exp\left(\frac{\tau_G^2}{2\tau_D^2} - \frac{t - t_0}{\tau_D}\right)\mathrm{erfc}\left(\frac{\tau_G^2 - \tau_D(t - t_0)}{\sqrt{2}\,\tau_D\tau_G}\right)$$

in which τ_D is the fluorescence lifetime of the free donor, τ_G is the width of the Guassian pulse response function, F_0 is the peak fluorescence before convolution and t_0 is the time offset, and erfc is the error function.

We fixed τ_D to the fluorescence lifetime obtained from free mEGFP (2.6 ns). To generate the fluorescence lifetime image, we calculated the mean photon arrival time, $\langle t \rangle$, in each pixel as:

$$\langle t \rangle = \int tF(t)dt \Big/ \int F(t)dt$$

then, the mean photon arrival time is related to the mean fluorescence lifetime, $\langle \tau \rangle$, by an offset arrival time, t_0, which is obtained by fitting the whole image:

$$\langle \tau \rangle = \langle t \rangle - t_0$$

For small regions-of-interest (ROIs) in an image (spines or dendrites), we calculated the binding fraction (P_{AD}) as:

$$P_{AD} = \tau_D(\tau_D - \langle \tau \rangle)(\tau_D - \tau_{AD})^{-1}(\tau_D + \tau_{AD} - \langle \tau \rangle)^{-1} \tag{3}$$

BDNF–SEP and BDNF–mRFP1 imaging. BDNF–SEP imaging was performed by interleaving 8 Hz two-photon imaging with two-photon glutamate uncaging (30 pulses at 0.5 Hz). Multiple (1–30) spines were imaged on each neuron. Change in BDNF–SEP fluorescence was measured as $\Delta F/F_0$ after subtracting background fluorescence. Uncaging-triggered averages were calculated as the average increase in SEP fluorescence after each individual uncaging pulse. Red fluorescence increase was smoothed using a 16-frame window.

For visualizing BDNF–mRFP1 localization in CA1 pyramidal neurons, images were obtained using a Leica SP5 laser scanning confocal microscope (Leica).

Spine volume analysis. During 2pFLIM and BDNF–SEP imaging (Figs 2, 3), spine volume was reported using the red fluorescent intensity from mRFP1 or mCh. For two-photon imaging without FLIM (Fig. 4), green fluorescent intensity from eGFP was used. In all experiments, spine volume was measured as the integrated fluorescent intensity after subtracting background (F). Spine volume change was calculated by F/F_0, in which F_0 is the average spine volume before stimulation. Additionally, to compare basal spine size/morphology between various conditions, maximal spine ($F_{\max(\text{spine})}$) and dendrite ($F_{\max(\text{dendrite})}$) fluorescent intensities were measured and the $F_{\max(\text{spine})}/F_{\max(\text{dendrite})}$ ratio was calculated after subtracting background fluorescence.

***In utero* viral injection for single-cell BDNF knockout.** E14.5/15.5 timed-pregnant *Bdnf*$^{fl/fl}$ mice were deeply anaesthetized using an isoflurane–oxygen mixture. The uterine horns were exposed and approximately 1–2 μl of AAV solution mix (containing AAV1.CAG.EGFP, AAV1.CAG.Flex.tdTomato and AAV1.hSyn.Cre, all from U Penn vector core) was injected through a pulled-glass capillary tube into the right lateral ventricle of each embryo. To achieve sufficient labelling of eGFP CA1 neurons alongside sparse expression of Cre + BDNF knockout tdTomato neurons, eGFP and Flex-tdTomato viruses were used at concentration of $\sim 10^{12}$ viral genome copies per μl, and Cre was diluted (∼100-fold) in PBS at a dilution determined to achieve a sparse labelling density of Cre-positive CA1 neurons.

Functional LTP. LTP experiments in Fig. 5 and Extended Data Fig. 10 were performed in Max Planck Florida Institute (MPFI) and Duke University, respectively. Mice (wild type, *Trkb*^{F616A}, or *Bdnf*$^{fl/fl}$ age 21–42 days) were sedated by isoflurane inhalation, and the brain was removed and dissected in a chilled cutting solution (124 mM choline chloride, 2.5 mM KCl, 26 mM $NaHCO_3$, 3.3 mM $MgCl_2$, 1.2 mM NaH_2PO_4, 10 mM D-glucose and 0.5 mM $CaCl_2$: MPFI or 110 mM sucrose, 60 mM NaCl, 3 mM KCl, 1.25 mM NaH_2PO_4, 28 mM $NaHCO_3$, 0.5 mM $CaCl_2$, 7.0 mM $MgCl_2$, and 5 mM D-glucose. The solutions were saturated with 95% O_2 plus 5% CO_2, pH 7.4)[37]. Coronal slices (250 μm: MPFI) or transverse hippocampal slices (400 μm: Duke) were prepared and maintained in oxygenated ACSF (MPFI/Duke: 127/124 mM NaCl, 2.5/1.75 mM KCl, 10/11 mM D-glucose, 26/25 mM $NaHCO_3$, 1.25/0 mM NaH_2PO_4, 0/1.25 mM KH_2PO_4, 1.3/2.0 $MgCl_2$ and 2.4/2.0 $CaCl_2$,) in a submerged chamber at 32–34 °C for at least 1 h before use.

Electrophysiological recordings were performed in ACSF (plus picrotoxin at MPFI). CA1 pyramidal neurons in acute hippocampal slices from wild-type and *Trkb*^{F616A} mice were visualized using oblique illumination or differential interference contrast (DIC). For *Bdnf*$^{fl/fl}$ experiments, Cre-negative (eGFP-expressing) and Cre-positive (tdTomato-expressing) neurons were identified and targeted with fluorescence microscopy. Patch pipettes (3–6 MΩ) were filled with an internal solution (130 mM K gluconate, 10 mM, Na phosphocreatine 4 mM $MgCl_2$,4 mM NaATP, 0.3 mM MgGTP, 3 mM L-ascorbic acid and 10 mM HEPES, pH 7.4, and 310 mOsm at MPFI or K-gluconate 140 mM, HEPES 10 mM, EGTA 1 mM, NaCl 4 mM, Mg_2ATP 4 mM, and Mg_2GTP 0.3 mM, pH 7.25, and 290 mOsm at Duke). Series resistances (10–40 MΩ) and input resistances (100–300 MΩ) were monitored throughout the experiment using negative voltage steps. The membrane potential was held at −70 mV. Experiments were performed at room temperature (∼21 °C) and slices were perfused with oxygenated ACSF. For *Trkb*^{F616A}/wild-type experiments, 1NMMP1 or vehicle was added to the ACSF before stimulation. For TrkB-Ig experiments, slices were incubated in 2 μg ml^{-1} TrkB-Ig or control human IgG for at least 2 h before the experiments. EPSCs were evoked by extracellular stimulation of Schaffer collaterals using a concentric bipolar stimulating electrode (World Precision Instruments) at a rate of 0.03 Hz. LTP was induced by pairing a 2-Hz stimulation with a postsynaptic depolarization to 0 mV for 15 s (MPFI) or 75 s (Duke). EPSC potentiation was assessed for 30–45 min (for *Trkb*^{F616A} experiments), 40–60 min (for *Bdnf*$^{fl/fl}$ experiments) or 20–30 min (for TrkB-Ig experiments) after stimulation.

Immunoprecipitation. HeLa cells were transfected with the TrkB sensor (TrkB–eGFP and mRFP1–PLC–mRFP1) using Lipofectamine 2000 as described above. Then, 24–48 h after transfection, the media bathing the cells was exchanged for HEPES buffered ACSF for biochemistry (150 mM NaCl, 3 mM KCl, 10 mM HEPES pH 7.35, 20 mM glucose, and 310 mOsM). After a 30-min equilibration period, cells were stimulated with 100 ng ml^{-1} BDNF for 10 min. Following stimulation, cells were washed in ice-cold PBS (Gibco), and then lysed in modified RIPA buffer (50 mM Tris-HCl pH 7.4, 150 mM NaCl, 1% NP-40, 0.25% sodium deoxycholate, 1 mM EDTA, 1 mM PMSF, 1 mM Na_3VO_4, and protease inhibitors) for 10 min on ice. The supernatant was collected after a 10 min centrifugation at 16,000*g* at 4 °C. At this point, a small volume of the supernatant was added to SDS-sample buffer and saved as the 'cell lysate' sample. The remaining supernatant was pre-cleared using protein G Sepharose beads (25 μl, Roche) for 30 min at 4 °C. After pre-clearing, the supernatant was incubated with 20 μg mouse monoclonal anti-phosphotyrosine (BD Transduction Labs) at 4 °C overnight. The immunocomplexes were precipitated with protein G Sepharose beads (50 μl) for 3 h at 4 °C and then analysed with western blotting. Antibodies used in western blotting included TrkB (Millipore), GFP (Abcam), actin (Sigma), and pTrkB(Y515) (Sigma).

Electron microscopic immunohistochemistry. Male adult (∼2–3 months old) *Bdnf-HA* knock-in mice[26] and aged matched wild-type C57/BL mice were used. The same investigator (T.A.M.) perfused all mice (*Bdnf-HA* and wild type) to maintain consistency between groups. Mice (3 per group) were deeply anaesthetized with sodium pentobarbital (150 mg kg^{-1}, i.p.) and perfused sequentially through the ascending aorta with: (1) ∼5 ml saline (0.9%) containing 2% heparin, and (2) 30 ml of 3.75% acrolein and 2% paraformaldehyde in 0.1 M phosphate buffer (PB; pH 7.4)[38]. Following removal from the skull, the brain was post-fixed for in 2% acrolein and 2% paraformaldehyde in PB 30 min. Brains were then sectioned (40 μm thick) on a Vibratome and stored at −20 °C in cryoprotectant until use.

For each animal, two dorsal hippocampal sections were processed for immunoelectron microscopy (immunoEM) experiments using previously described methods[38]. Before immunohistochemical processing, sections were rinsed in PB, and experimental groups were coded with hole-punches so that tissue could be run in single crucibles, ensuring identical exposure to all reagents.

Before processing for immunolabelling, sections were treated with 1% sodium borohydride for 30 min to remove free aldehyde sites. Sections then were rinsed in PB followed by a rinse in 0.1 M Tris-saline (TS; pH 7.6) and then a 30 min incubation in 0.5% BSA in TS. Sections then were incubated in primary rabbit anti-HA (1:1,000; Sigma) in 0.025% Triton-X 100 and 0.1% BSA in TS for 1 day at room temperature and 4 days at 4 °C. Sections then were incubated in donkey anti-rabbit biotinylated IgG (1:400; Jackson Immunoresearch Laboratories) for 30 min followed by a 30 min incubation in avidin-biotin complex (ABC; Vectastain Elite Kit, Vector Laboratories) in TS (1:100 dilution). Sections were developed in 3,3′-diaminobenzidine (Sigma-Aldrich) and H_2O_2 in TS. All antibody incubations were performed in 0.1% BSA/TS and separated by washes in TS.

Sections were post-fixed in 2% osmium tetroxide for 1 h, dehydrated, and flat embedded in Embed-812 (EMS) between two sheets of Aclar plastic. Brain sections containing the CA1 and dentate gyrus were selected from the plastic embedded sections, glued onto Epon blocks and trimmed to 1 mm-wide trapezoids. Ultra-thin sections (70 nm thickness) through the tissue-plastic interface were cut with a diamond knife (EMS) on a Leica EM UC6 ultratome, and sections were collected on 400-mesh, thin-bar copper grids (EMS). Grids were then counterstained with uranyl acetate and Reynold's lead citrate.

Ultrastructural analysis. An investigator blinded to animal condition performed the data collection and analysis. One section from each of *Bdnf-HA* and wild-type animals was analysed ($n = 3$ each group). The thin sections were examined and photographed on a Tecnai Biotwin transmission electron microscope (FEI). Cell profiles were identified by defined morphological criteria[39]. Dendritic profiles generally were postsynaptic to axon terminals and contained regular microtubule arrays. Dendritic spines also were usually postsynaptic to axon terminal profiles and sometimes contained a spine apparatus. Axon terminals contained small synaptic vesicles and occasional dense-core vesicles. Unmyelinated axons were profiles smaller than 0.15 μm that contained a few small synaptic vesicles and lacked a synaptic junction in the plane of section. Glial profiles were distinguished by the presence of glial filaments (astrocytic profiles), by the presence of microtubules and/or their tendency to conform irregularly to the boundaries of surrounding profiles. 'Unknown profiles' were those that contained immunoperoxidase reaction product but could not be definitively placed in one of the above categories.

From each block, 4 grid squares (each square was 55 × 55 μm^2) each from the CA1 near stratum radiatum (nSR in Fig. 3; that is, adjacent to the pyramidal cell layer) and distal stratum radiatum (dSR in Fig. 3; that is, 50–150 μm away from the pyramidal cell layer) were randomly sampled for analysis. Thus, 12,100 μm^2 was sampled for each area in each block. Grid squares were selected plastic-tissue interface to ensure even antibody tissue penetration[38]. Immunoperoxidase labelling for HA was evident as a characteristic, electron-dense DAB reaction product precipitate. All peroxidase labelled profiles from each square were photographed and categorized. Animal codes were not broken until all 6 blocks were analysed.

Statistical analysis. Sample sizes for all experiments were chosen based on signal-to-noise ratios identified in pilot experiments. Variances of all data sets were estimated and compared using Bartlett's or Levene's test before further statistical analysis. Randomization of animals and/or slices was not needed.

To evaluate distribution patterns of TrkB sensor activity, spine volume change, and BDNF–SEP signal, peak responses for each data set (the same points used for statistical comparisons) were subjected to a Shapiro–Wilk test for normality. TrkB sensor activity adhered to the null hypothesis (normal distribution) while spine volume change and BDNF–SEP signal did not.

Because TrkB sensor activity had a normal distribution, parametric statistics were used: paired and unpaired two-tailed *t*-test, ANOVA, and repeated-measures ANOVA with appropriate post-hoc analysis, as indicated in the figure legends and supplementary note. For *t*-tests, homoscedasticity between groups was evaluated

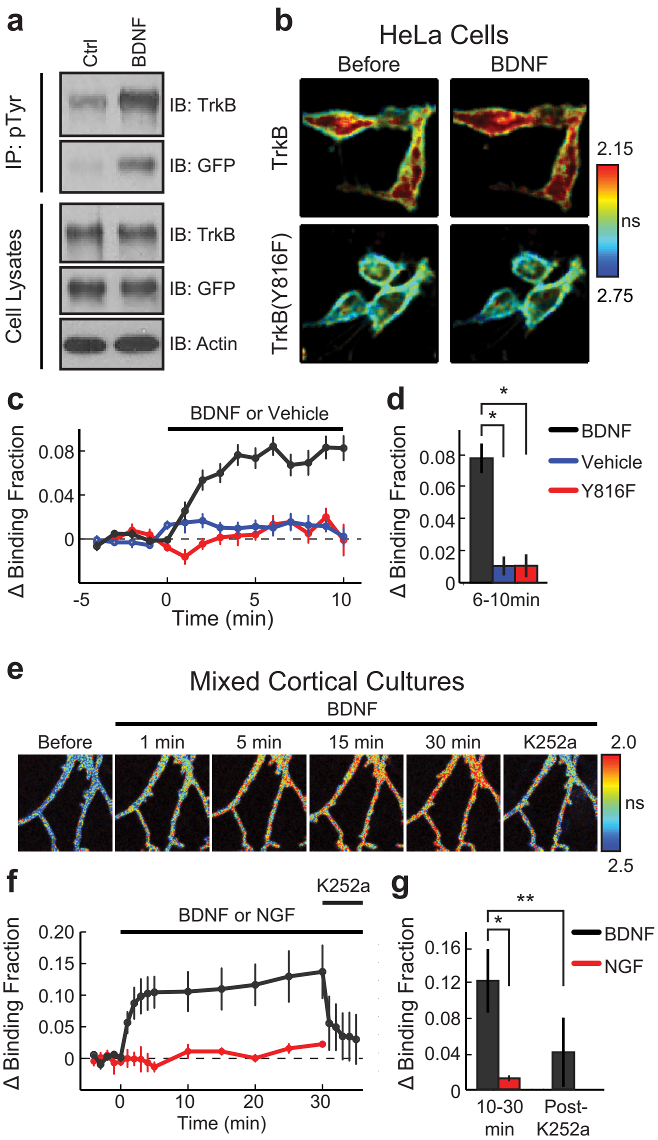
using the F-test. If variance was unequal, Welch's corrected t-test was performed. For ANOVA, homoscedasticity was evaluated with Bartlett's test. For multiple comparisons of sensor activity, data were subjected to ANOVA or repeated-measures ANOVA followed by a post-hoc test to determine statistical significance. In cases where each condition was compared to all other conditions in the experiment, the Tukey–Kramer method was employed. In cases where each condition was compared to a single control, Dunnet's test was used.

Since spine volume change had a non-normal distribution, data were log-transformed to resolve skewness and then analysed with parametric statistics (the same tests described above), as indicated in the figure legends.

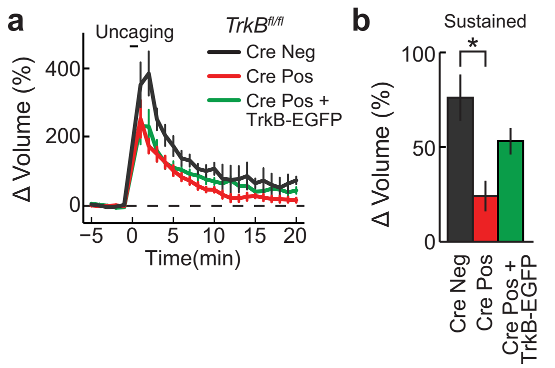
For the BDNF–SEP signal, log-transformation of the data did not resolve the skewness. As such, non-parametric statistics were used—Wilcoxon rank-sum test and Kruskal–Wallis test with followed by a Dunn's test.

Data were only excluded if obvious signs of poor cellular health (dendritic blebbing, spine collapse, etc.) were apparent.

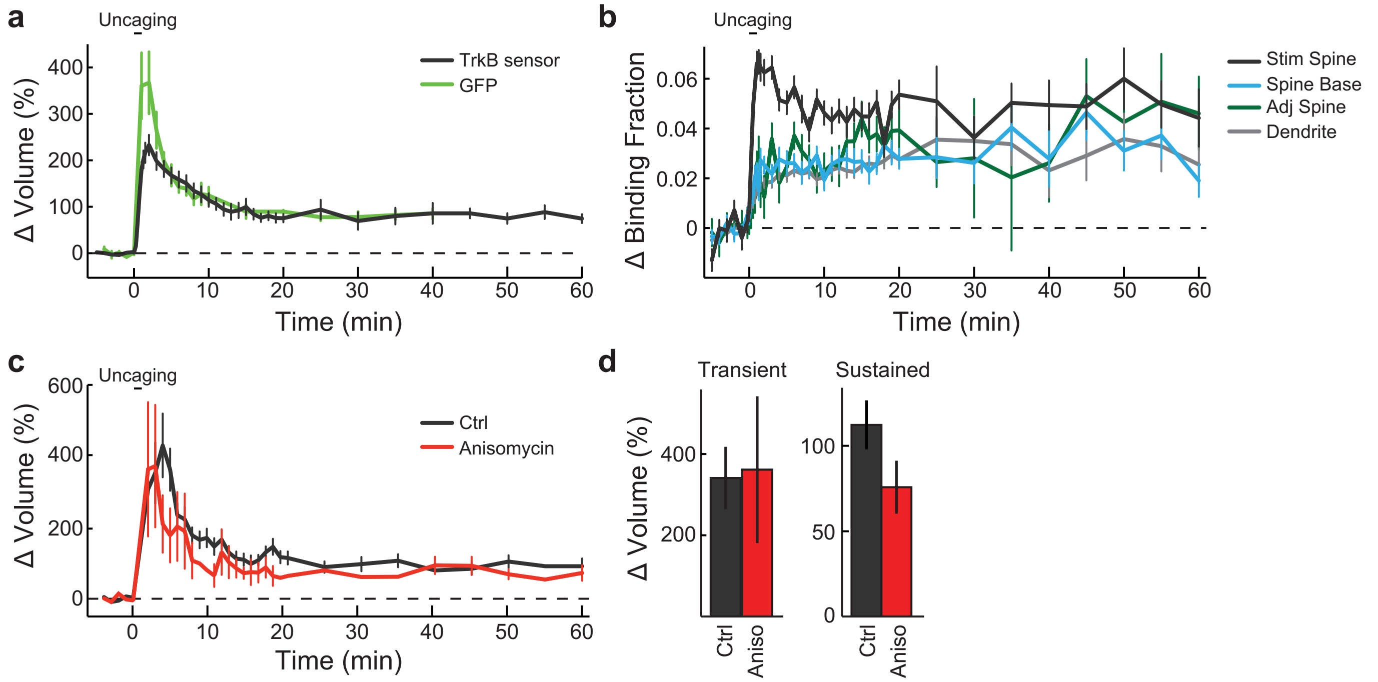
29. Huang, Y. Z. & McNamara, J. O. Mutual regulation of Src family kinases and the neurotrophin receptor TrkB. *J. Biol. Chem.* **285,** 8207–8217 (2010).
30. Zacharias, D. A., Violin, J. D., Newton, A. C. & Tsien, R. Y. Partitioning of lipid-modified monomeric GFPs into membrane microdomains of live cells. *Science* **296,** 913–916 (2002).
31. Patterson, M. A., Szatmari, E. M. & Yasuda, R. AMPA receptors are exocytosed in stimulated spines and adjacent dendrites in a Ras-ERK-dependent manner during long-term potentiation. *Proc. Natl Acad. Sci. USA* **107,** 15951–15956 (2010).
32. He, X.-P. P. *et al.* Conditional deletion of TrkB but not BDNF prevents epileptogenesis in the kindling model. *Neuron* **43,** 31–42 (2004).
33. Xiong, Z. Q. & McNamara, J. O. Fleeting activation of ionotropic glutamate receptors sensitizes cortical neurons to complement attack. *Neuron* **36,** 363–374 (2002).
34. Stoppini, L., Buchs, P. A. & Muller, D. A simple method for organotypic cultures of nervous tissue. *J. Neurosci. Methods* **37,** 173–182 (1991).
35. Murakoshi, H., Lee, S.-J. J. & Yasuda, R. Highly sensitive and quantitative FRET-FLIM imaging in single dendritic spines using improved non-radiative YFP. *Brain Cell Biol.* **36,** 31–42 (2008).
36. Pologruto, T. A., Sabatini, B. L. & Svoboda, K. ScanImage: flexible software for operating laser scanning microscopes. *Biomed. Eng. Online* **2,** 13 (2003).
37. Pan, E. *et al.* Vesicular zinc promotes presynaptic and inhibits postsynaptic long-term potentiation of mossy fiber-CA3 synapse. *Neuron* **71,** 1116–1126 (2011).
38. Milner, T. A., Waters, E. M., Robinson, D. C. & Pierce, J. P. in *Neurodegeneration, Methods and Procedures* (eds Manfredi, G. & Kawamata, H.) 23–59 (Spring, 2011).
39. Peters, A., Palay, S. L. & Webster, H. D. *The Fine Structure of the Nervous System* 3rd edn (Oxford Univ. Press, 1991).



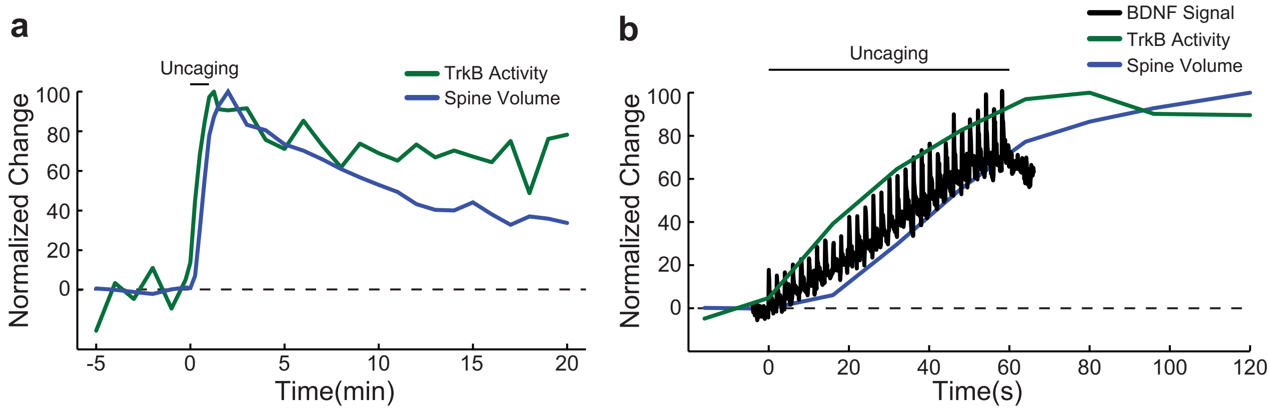
Extended Data Figure 1 | Design and development of a FRET-based sensor for TrkB activation. a, Top, western blot analysis of cell extracts from HeLa cells stimulated with either BDNF or vehicle. Extracts were immunoprecipitated with an antibody for phosphorylated tyrosine residues (pTyr) and then probed with antibodies for TrkB and GFP. Bottom, immunoblot (IB) of BDNF and vehicle stimulated cell extracts before immunoprecipitation (IP) using antibodies for TrkB, GFP and actin. For source data, see Supplementary Fig. 1. **b**, FLIM images of TrkB and $\mathrm{TrkB}^{\mathrm{Y816F}}$ activation acquired before and 2–6 min after BDNF stimulation (averaged multiple images taken over 5 min). Warmer colours indicate shorter lifetimes and higher TrkB activity. **c**, Time course of TrkB and $\mathrm{TrkB}^{\mathrm{Y816F}}$ activation measured as the change in binding fraction of TrkB–eGFP or $\mathrm{TrkB}^{\mathrm{Y816F}}$–eGFP bound to mRFP1–PLC–mRFP1 before and after BDNF or vehicle stimulation. $n = 22/8$ TrkB plus BDNF, 9/4 TrkB plus vehicle, and 11/4 $\mathrm{TrkB}^{\mathrm{Y816F}}$ plus BDNF (cells/experiments). **d**, TrkB activation (averaged over 6–10 min) for experiments in **c**. **e**, FLIM images of TrkB activation in a neuron in a mixed cortical dissociated culture before and after BDNF stimulation followed by K252a application at 30 min. **f**, Time course of TrkB activation measured as described in **c** before and after BDNF or NGF stimulation followed by K252a application. $n = 8$ BDNF and 4 NGF (neurons). **g**, TrkB activation (averaged over 10–30 min and 3–5 min following K252a application) for experiments in **f**. Data are mean ± s.e.m. $^*P < 0.05$ as determined by a two-tailed unpaired samples *t*-test (**g**) or an analysis of variance (ANOVA) followed by Tukey's method to correct for multiple comparisons. (**d**). $^{**}P < 0.05$ as determined by a two-tailed paired samples *t*-test.



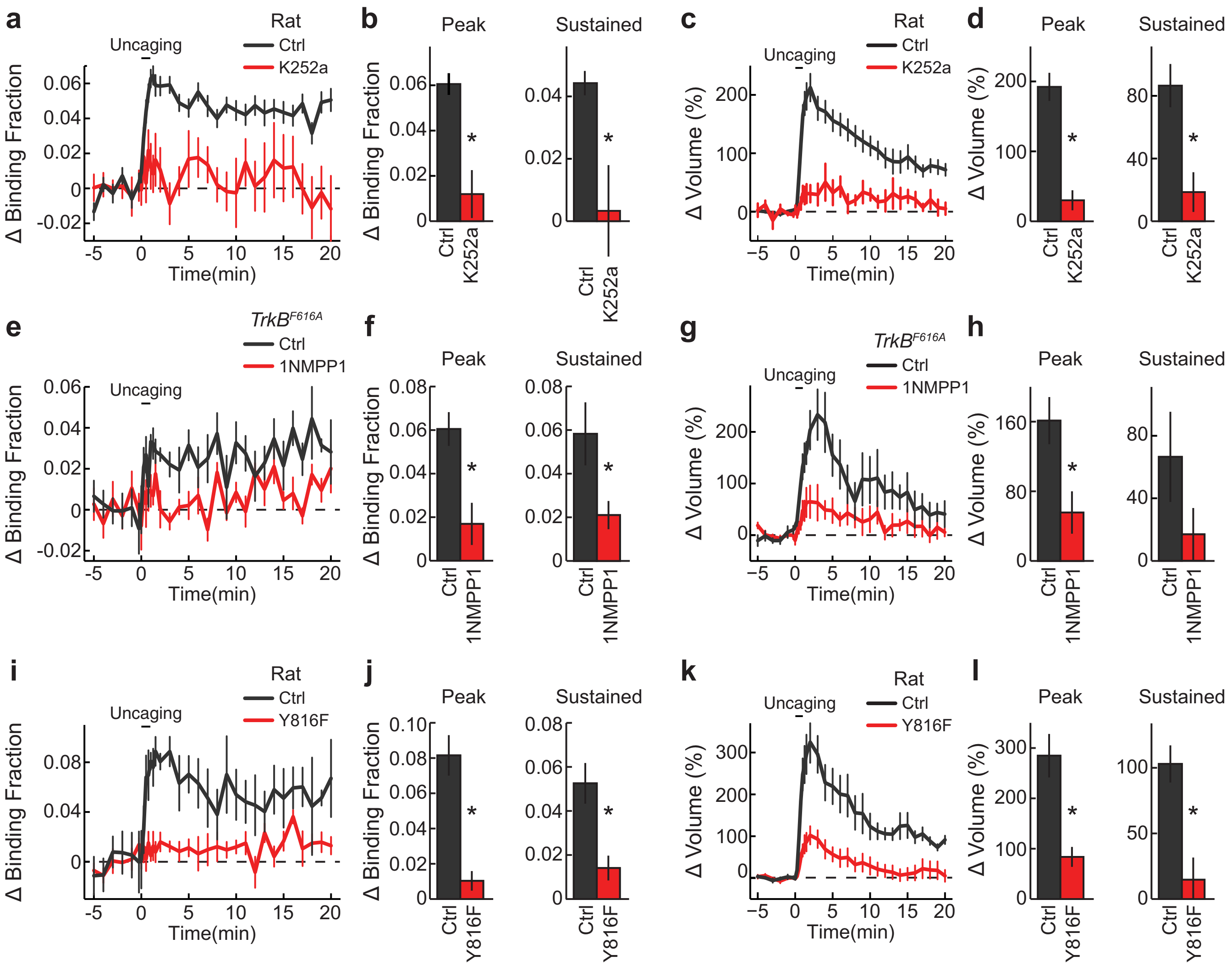
Extended Data Figure 2 | Rescue of sLTP with TrkB–eGFP following postsynaptic TrkB knockout. a, **b**, Time course (**a**) and quantification (**b**) of glutamate-uncaging-induced spine volume change for *Trkb*$^{fl/fl}$ hippocampal slices transfected with eGFP (Cre Neg), eGFP plus Cre (Cre Pos), and mCh, TrkB–eGFP and Cre (Cre Pos + TrkB–eGFP). $n = 7/20$ Cre Neg, 9/24 Cre Pos, and 5/11 Cre Pos + TrkB–eGFP (cells/spines). Data are mean ± s.e.m. $^*P < 0.05$ as determined by an ANOVA followed by Tukey's method to correct for multiple comparisons.



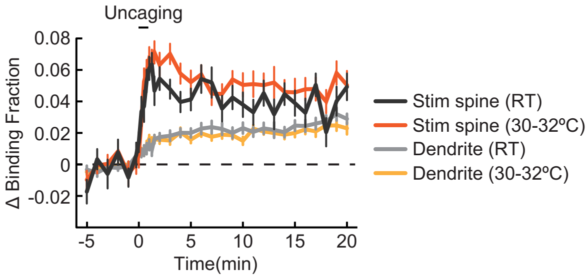
Extended Data Figure 3 | Characterization of prolonged TrkB activation and spine volume change during single spine sLTP. **a**, Prolonged time course of spine volume change after two-photon glutamate uncaging in rat hippocampal slices transfected with the TrkB sensor or eGFP. $n = 50/54$ for TrkB sensor (9/10 for experiments longer than 20 min) and 8/8 for eGFP (cells/spines). **b**, Prolonged time course of TrkB activation in stimulated spines, the base of the spine neck, adjacent spines, and the dendritic shaft adjacent to the stimulated spine. $n = 50$ cells with 54 stimulated spine, spine base, and dendrite plus 59 adjacent spine. **c**, **d**, Time course (**c**) and quantification (**d**) of the transient (averaged over 1–2 min) and sustained (averaged over 20–40 min) phases of glutamate uncaging-induced spine volume change in rat hippocampal slices in the absence and presence of anisomycin (25 μM). $n = 12/14$ control and 5/5 anisomycin (cells/spines). Data are mean ± s.e.m.



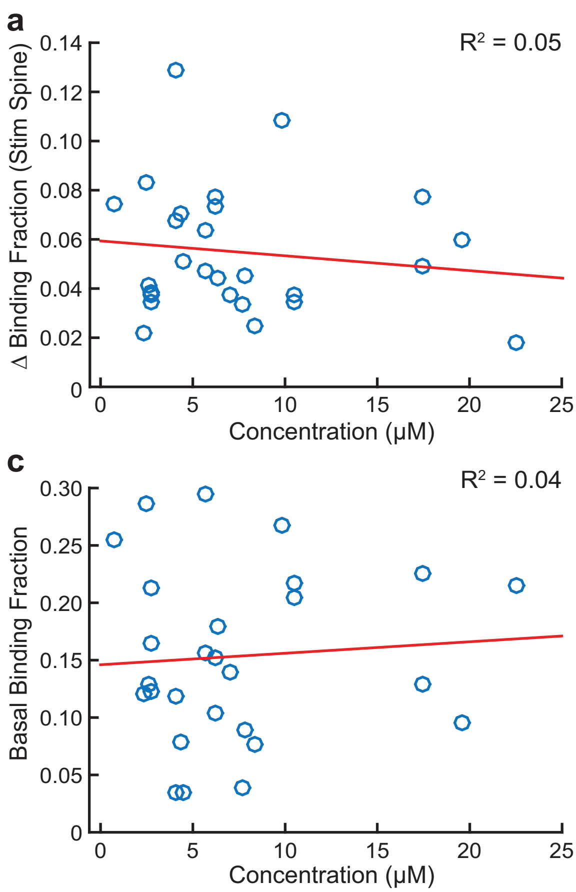
Extended Data Figure 4 | Comparison of temporal dynamics of BDNF release, TrkB activation, and spine volume change during single spine sLTP. a, Time course of normalized changes in TrkB activity and spine volume change (percentage of maximal activity and volume change). **b**, Magnified view of normalized changes of BDNF release, TrkB activation, and spine volume during and 1 min after the uncaging epoch.



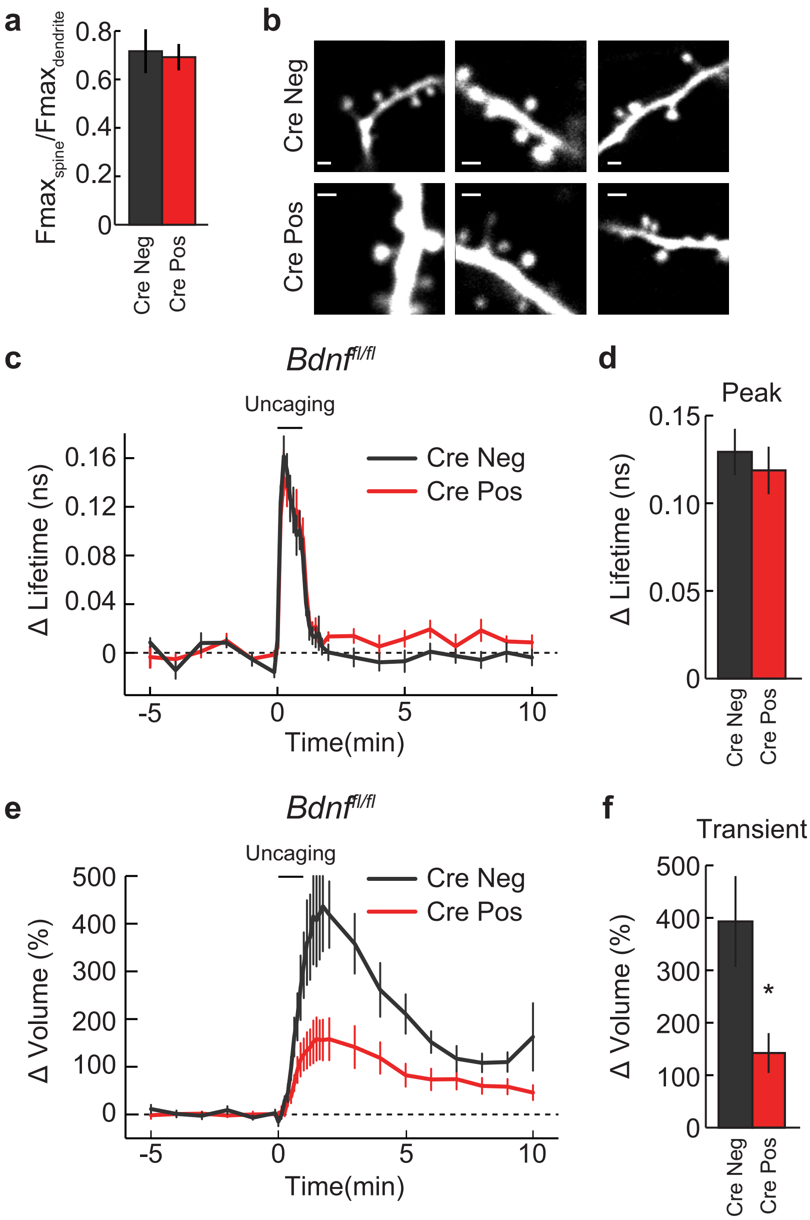
Extended Data Figure 5 | Determination of the specificity of glutamate uncaging evoked TrkB activation. a, Time course of TrkB activation following glutamate uncaging before and at least 30 min after K252a application to the perfusion bath. $n = 41/45$ Ctrl and 4/9 K252a (cells/spines). **b**, Peak (averaged over 1–2 min) and sustained (averaged over 10–20 min) TrkB activation for experiments in **a**. **c**, Time course of spine volume change for experiments in **a**. **d**, Transient and sustained spine volume change for experiments in **a**. **e–h**, Similar experiments to **a–d** but in *Trkb*^{F616A} hippocampal slices transfected with the TrkB^{F616A} sensor before and at least 30 min after 1NMPP1 application (2 μM). $n = 4/5$ control and 3/6 1NMPP1 (cells/spines). **i–l**, Similar experiments to **a–d** but with the TrkB and TrkB^{Y816F} sensors. $n = 9/10$ control and 7/11 Y816F (cells/spines). Data are mean ± s.e.m. $*P < 0.05$ as determined by two-tailed unpaired samples *t*-test.



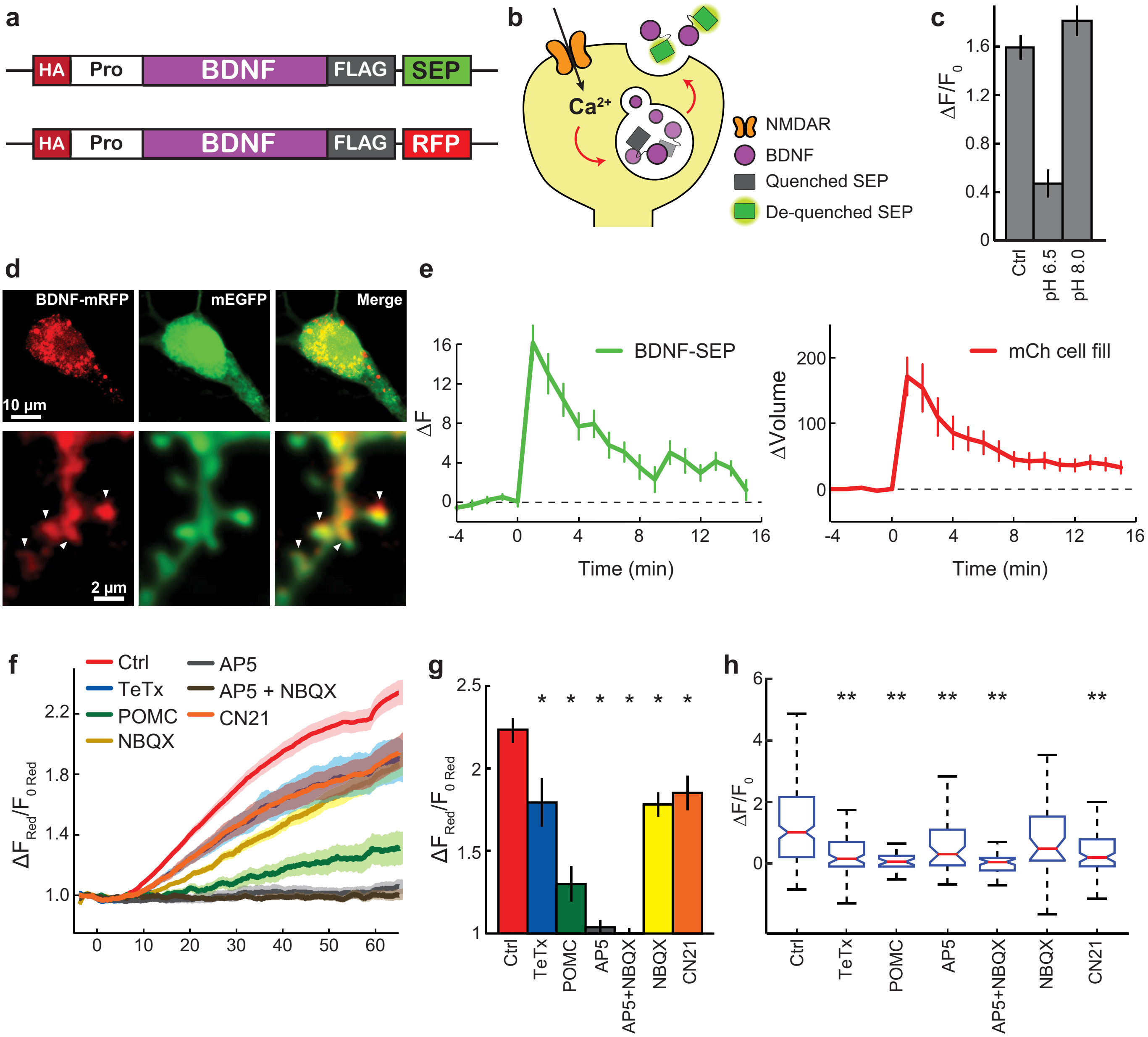
Extended Data Figure 6 | Effect of temperature on the spatiotemporal dynamics of TrkB activation. Time course of TrkB activation at room temperature (RT; 24–26 °C) and 30–32 °C in the stimulated and dendrite. $n = 19/20$ and 23/25 at 24–26 °C and 30–32 °C, respectively (spines/cells). Data are mean ± s.e.m.



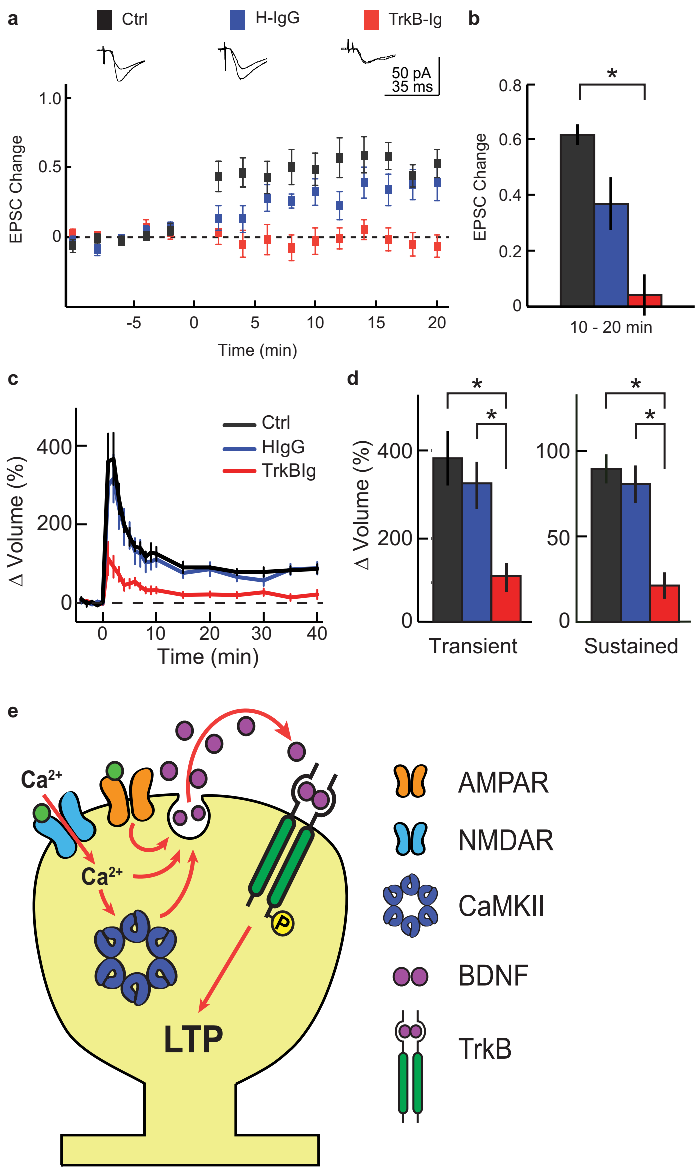
Extended Data Figure 7 | Effects of sensor expression levels on changes reported by the sensor. a–d, Effect of TrkB–eGFP concentration as measured in individual neurons on corresponding change in binding fraction of the stimulated spine (**a**), change in spine volume (**b**), binding fraction before uncaging (basal binding fraction) (**c**), and change in binding fraction of the dendrite (**d**). $n = 25/28$ (cells/spines). Data are mean values and were fit to a linear regression model with corresponding coefficients of determination (R^2) provided for each.



Extended Data Figure 8 | Basal spine size and CaMKII activation in the presence and absence of post- synaptic BDNF. a, b, Quantification (**a**) and representative two-photon images (**f**) of basal spine size/morphology in *Bdnf*$^{fl/fl}$ slices transfected with eGFP or eGFP plus Cre (Cre Neg or Pos). $n = 14/50$ Cre Neg and 29/117 Cre Pos (cells/spines). Scale bar, 1 μm. **c, d,** Time course (**c**) and quantification (averaged over 0–45 s) (**d**) of CaMKII activation in *Bdnf*$^{fl/fl}$ slices transfected with the CaMKII sensor or CaMKII plus Cre. $n = 7/13$ Cre Neg and 7/15 for Cre Pos (cells/spines). **e, f,** Time course and quantification of the transient phase of spine volume change for experiments in **c.** Data are mean ± s.e.m. $^*P < 0.05$ as determined by a two-tailed unpaired samples *t*-test.



Extended Data Figure 9 | Design and validation of BDNF–SEP. **a**, Schematic of BDNF–SEP and BDNF–mRFP1. Pro, amino acids 19–128 of human BDNF; BDNF, amino acids 129–247 of human BDNF corresponding to the mature chain. **b**, Mechanistic model linking changes in SEP fluorescence with BDNF release. **c**, Change in BDNF–SEP fluorescence following glutamate uncaging under control, acidic (pH 6.5), and basic (pH 8.0) conditions. **d**, Confocal images of a CA1 pyramidal neuron transfected with eGFP and BDNF–mRFP1. Arrowheads indicate dendritic spines. **e**, Prolonged time course of BDNF–SEP fluorescence change (left) and spine volume change (right) in response to glutamate uncaging. $n = 11/20$ (cells/spines). **f**, **g**, Time course (**f**) and quantification (**g**) of spine volume change for experiments in Fig. 4c, d. $n = 31/218$ control, 6/82 TeTx, 2/29 POMC, 3/50 AP5, 2/46 AP5 + NBQX, 4/40 NBQX, and 7/88 CN21 (cells/spines). **h**, Data from Fig. 4d presented as median ± interquartile range. Data are mean ± s.e.m. unless otherwise indicated. $*P < 0.05$ as determined by an ANOVA followed by Dunnet's method to correct for multiple comparisons. $**P < 0.05$ as determined by a Kruskal–Wallis test followed by Dunn's test.



Extended Data Figure 10 | CA1-LTP requires exogenous BDNF. **a**, Time course of average EPSC amplitude changes recorded in CA1 pyramidal cells evoked by Schaffer collateral stimulation before and after LTP induction in the absence or presence of human-IgG or TrkB-Ig. Representative traces are above the graphs. $n = 22$ control, 9 HIgG, and 12 TrkB-Ig (animals). **b**, Quantification of EPSC amplitude changes averaged over 10–20 min following LTP induction. **c**, **d**, Time course (**c**) and quantification (**d**) of the transient and sustained glutamate-uncaging-induced spine volume change in rat hippocampal slices in the absence or presence of human-IgG or TrkB-Ig. $n = 8/8$ control, 6/8 TrkB-Ig, and 4/6 HIgG (cells/spines). **e**, Model of spine autonomous, autocrine, BDNF release and postsynaptic TrkB activation. Data are mean ± s.e.m. $^{*}P < 0.05$ as determined by an ANOVA followed by Tukey's method to correct for multiple comparisons

doi:10.1038/nature19784

Rho GTPase complementation underlies BDNF-dependent homo- and heterosynaptic plasticity

Nathan G. Hedrick[1]*†, Stephen C. Harward[1]*, Charles E. Hall[1], Hideji Murakoshi[2], James O. McNamara[1] & Ryohei Yasuda[1,3]

The Rho GTPase proteins Rac1, RhoA and Cdc42 have a central role in regulating the actin cytoskeleton in dendritic spines[1], thereby exerting control over the structural and functional plasticity of spines[2–5] and, ultimately, learning and memory[6–8]. Although previous work has shown that precise spatiotemporal coordination of these GTPases is crucial for some forms of cell morphogenesis[9], the nature of such coordination during structural spine plasticity is unclear. Here we describe a three-molecule model of structural long-term potentiation (sLTP) of murine dendritic spines, implicating the localized, coincident activation of Rac1, RhoA and Cdc42 as a causal signal of sLTP. This model posits that complete tripartite signal overlap in spines confers sLTP, but that partial overlap primes spines for structural plasticity. By monitoring the spatiotemporal activation patterns of these GTPases during sLTP, we find that such spatiotemporal signal complementation simultaneously explains three integral features of plasticity: the facilitation of plasticity by brain-derived neurotrophic factor (BDNF), the postsynaptic source of which activates Cdc42 and Rac1, but not RhoA; heterosynaptic facilitation of sLTP, which is conveyed by diffusive Rac1 and RhoA activity; and input specificity, which is afforded by spine-restricted Cdc42 activity. Thus, we present a form of biochemical computation in dendrites involving the controlled complementation of three molecules that simultaneously ensures signal specificity and primes the system for plasticity.

Previous studies using two-photon fluorescence lifetime imaging (2pFLIM) in combination with fluorescence resonance energy transfer (FRET)-based biosensors revealed that the Rho GTPases Cdc42 and RhoA had distinct spatial profiles during sLTP, with Cdc42 showing synapse-restricted activity and RhoA showing a diffuse, heterosynaptic pattern[3]. To obtain a more complete understanding of the spatiotemporal patterning of Rho GTPase activity during sLTP, we developed a FRET-based sensor for Rac1, following the design of the Cdc42 and RhoA sensors[3] (Fig. 1a; validation in Extended Data Figs 1–3). We transfected rat organotypic hippocampal slices with the sensor using biolistics[10,11] and imaged CA1 pyramidal neurons using 2pFLIM[12]. When sLTP was induced in single dendritic spines with two-photon glutamate uncaging[13,14] (Fig. 1b, c), Rac1 was rapidly (within ∼1 min) activated in the stimulated spine, and remained active for at least 30 min (Fig. 1b, d), notably displaying a more pronounced sustained phase than RhoA or Cdc42 (ref. 3). This activation initially showed limited diffusion, after which it slowly spread over ∼10 μm of the parent dendrite until it nearly equalized with activity in the stimulated spine (black), nearby spines (green), dendrite near the stimulated spine (cyan) and whole dendrite in the image (grey).

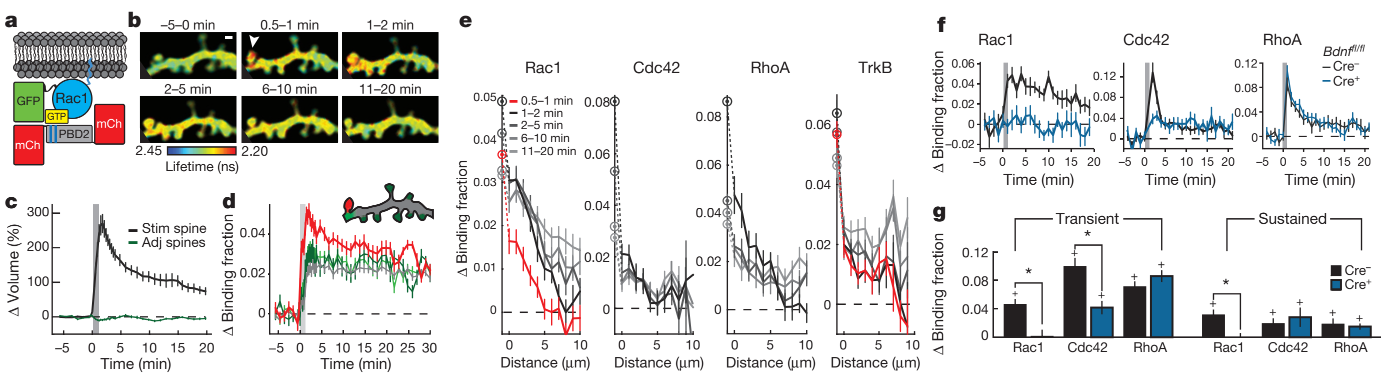


Figure 1 | The Rho GTPases Rac1 and Cdc42 convey postsynaptic BDNF–TrkB signalling across both homosynaptic and heterosynaptic domains. a, Schematic of Rac1 sensor. Monomeric enhanced green fluorescent protein (eGFP) is N-terminally tagged to Rac1 to preserve C-terminal membrane association. GTP binding leads to association with the Pak GTPase binding domain of PAK2^{R71C,S78A} (PBD2; blue lines denote mutations), bringing the mCherry (mCh) fluorophores within the FRET distance of eGFP, decreasing its fluorescence lifetime. **b**, Representative 2pFLIM images of Rac1 activation in dendrites during sLTP induced in a single spine with two-photon glutamate uncaging (white arrowhead). Scale bar, 1 μm. **c**, Time course of spine volume changes during sLTP induced with two-photon glutamate uncaging (grey window) in the stimulated spine (black) and compared to adjacent spines (green). $n = 102$ cells/121 spines. **d**, Time course of Rac1 activation during sLTP, measured as a change in the fraction of acceptor-bound eGFP–Rac1 in the stimulated spine (black), nearby spines (green), dendrite near the stimulated spine (cyan) and whole dendrite in the image (grey). $n = 102/121$ (cells/spines). **e**, Activations of Rac1 ($n = 56/79$), Cdc42 ($n = 25/38$), RhoA ($n = 21/23$), and TrkB ($n = 48/52$) (cells/spines) in dendrites as a function of distances from the base of the stimulated spines (lines) and in the stimulated spines (circles). **f**, Dependence of Rho GTPase activation on postsynaptically synthesized BDNF. Blue indicates co-expression of Cre recombinase in *Bdnf*$^{fl/fl}$ slices along with the Rac1 (left; $n = 7/13$ Cre$^-$, $n = 8/16$ Cre$^+$), Cdc42 (middle; $n = 5/12$ Cre$^-$, $n = 4/11$ Cre$^+$), or RhoA (right; $n = 6/13$ Cre$^-$, $n = 7/14$ Cre$^+$) (cells/spines) sensor. Black represents the corresponding control (Cre$^-$) data. **g**, Summary of data from **f**. Bars represent the average of the activation 1–2 min after stimulation. Error bars represent s.e.m. *$P < 0.05$ (two-tailed t-test between groups); $+P < 0.05$ (t-test compared to the baseline).

[1]Neurobiology Department, Duke University Medical Center, Research Drive, Durham, North Carolina 27710, USA. [2]National Institute for Physiological Science, Myodaiji, Okazaki, Aichi, 444-8585, Japan. [3]Max Planck Florida Institute for Neuroscience, 1 Max Planck Way, Jupiter, Florida 33458, USA. †Present address: Neurobiology Section, Center for Neural Circuits and Behavior, and Department of Neurosciences, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093, USA.
*These authors contributed equally to this work.

in the spine (Fig. 1b, d, e, Extended Data Fig. 2a, b). Similar to Cdc42 and RhoA, Rac1 activation was dependent on NMDARs (*N*-methyl-D-aspartate receptors) and CaMKII (Extended Data Fig. 2d–f), and both pharmacological inhibition of Rac1 and single-cell knockout of Rac1 inhibited sLTP (Extended Data Fig. 4). Thus, like RhoA and Cdc42, Rac1 is a Rho GTPase molecule linking NMDAR–CaMKII signalling to sLTP.

We next examined whether postsynaptic BDNF[15], which also links NMDAR–CaMKII signalling and sLTP, is required for Rho GTPase activation during sLTP. To do this, we used a single-cell knockout technique[15,16] with organotypic hippocampal slices from BDNF conditional knockout mice (*Bdnf*$^{fl/fl}$)[17,36] coupled with biolistic transfection of Cre recombinase alongside a Rho GTPase sensor. We found that removal of postsynaptic BDNF significantly attenuated Rac1 and Cdc42 activation during sLTP without affecting RhoA (Fig. 1f, g), and reduced the associated expression of sLTP[15] (Extended Data Fig. 5a, b). This suggests that Rac1 and Cdc42, but not RhoA, are downstream of postsynaptic BDNF. Similarly, addition of the extracellular BDNF scavenger TrkB-Ig (2 μg ml^{-1}) significantly attenuated the activation of both Rac1 and Cdc42 (Extended Data Fig. 5c, d, g, h). Finally, postsynaptic removal of the BDNF receptor TrkB using *Trkb*$^{fl/fl}$ (also known as *Ntrk2*$^{fl/fl}$) mice also significantly attenuated Rac1 and Cdc42 activation (Extended Data Fig. 5e, f, i, j). These results suggest that an autocrine BDNF–TrkB system controls the activation of Rac1 and Cdc42, thereby instructing two distinct spatial signalling domains relevant to plasticity: a spine-specific domain comprising BDNF–TrkB–Cdc42, and a diffuse domain comprising BDNF–TrkB–Rac1. Since BDNF–TrkB signalling is preferentially enriched in stimulated spines during sLTP (Fig. 1e, Extended Data Fig. 2a), diffusion of Rac1 molecules (Extended Data Fig. 2c) or of Rac1 activators downstream of TrkB[7,18] probably caused the observed spreading of Rac1 activity. By contrast, the spatial spreading of Cdc42 (before its decay) was similar to that of TrkB during the same time period (Extended Data Fig. 2b).

Notably, the length scale of Rac1 spreading is similar to that of synaptic crosstalk, a phenomenon in which sLTP induction briefly facilitates sLTP in nearby (∼5–10 μm) spines on the same dendrite[19,20]. Thus, we next addressed whether BDNF–TrkB–Rac1 signalling contributes to such crosstalk. Consistent with previous studies[19,20], a suprathreshold sLTP stimulus delivered to one spine allowed a subsequent subthreshold stimulus at a nearby (≤∼5 μm) spine to induce sLTP (hereafter referred to as crosstalk) (Fig. 2a, b). To test whether crosstalk requires BDNF–TrkB–Rac1 signalling, we first used partial pharmacological inhibition of this pathway. Although strong inhibition of BDNF signalling has been shown to impair sLTP[15], we found that weak inhibition of this pathway with a low concentration (0.25 μg ml^{-1}) of TrkB-Ig preserved sLTP ($\Delta V_{\text{supra sustained}} = 58 \pm 13\%$ (mean ± s.e.m.), in which V denotes spine volume), but attenuated crosstalk ($\Delta V_{\text{sub sustained}} = 20 \pm 8\%$), suggesting that BDNF is required for this process (Fig. 2c, g). To ensure that these effects were due to BDNF signalling through TrkB, we used *Trkb*^{F616A} mutant mice, which contain a point mutation that renders TrkB uniquely susceptible to inhibition by the small molecule 1NMPP1 (ref. 21). We found that application of a low concentration (0.125 μM) of 1NMPP1 in *Trkb*^{F616A} slices selectively abolished crosstalk without affecting sLTP ($\Delta V_{\text{supra sustained}} = 59 \pm 16\%$, $\Delta V_{\text{sub sustained}} = 9 \pm 8\%$) (Fig. 2d, g). Interestingly, weak BDNF–TrkB inhibition also reduced Rac1 activity in dendritic shafts during the time frame of crosstalk induction (1–2 min), without inhibiting the activity

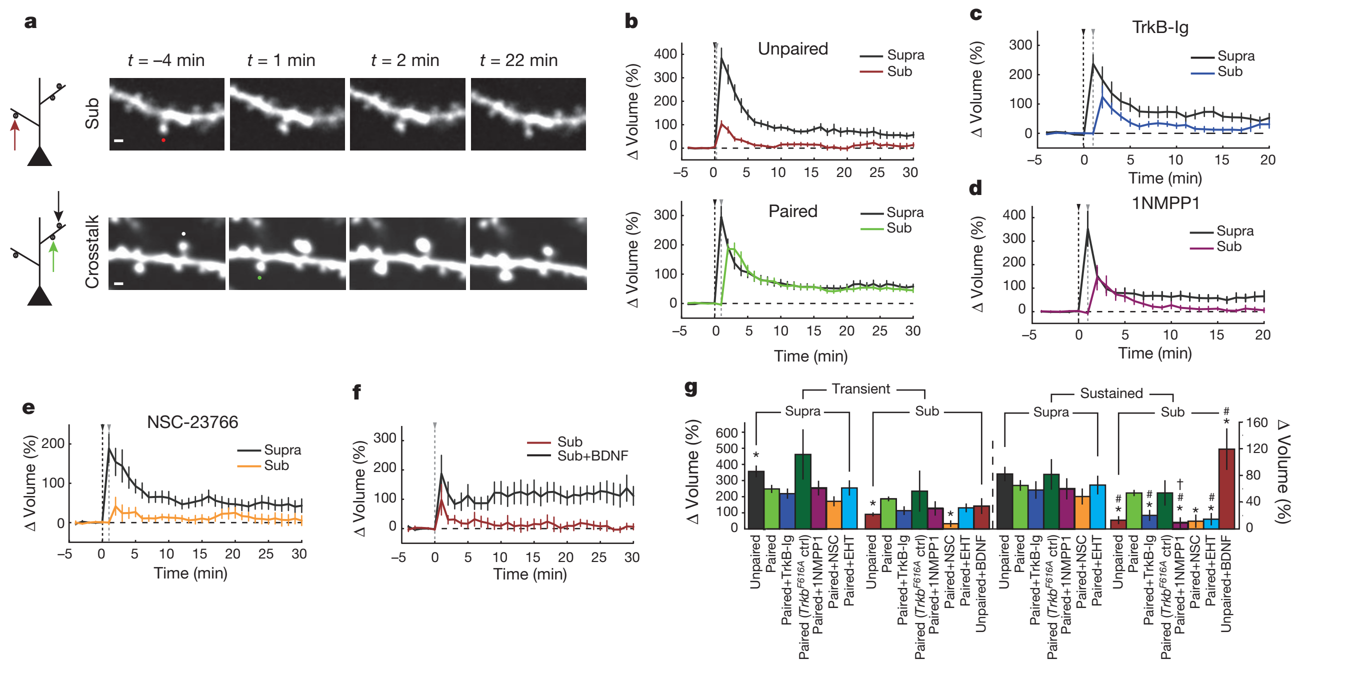


Figure 2 | BDNF–TrkB–Rac1 signalling is required for synaptic crosstalk. a, Left, schematic of crosstalk model. Top, in 'unpaired' trials, the subthreshold (sub) stimulus was delivered to a single spine. Bottom, in the 'paired' (crosstalk) condition, a threshold stimulus was delivered to a spine before the delivery of a subthreshold stimulus to a nearby spine on the same dendrite. Right, representative images of unpaired (top) and paired (bottom) subthreshold stimuli. Scale bars, 1 μm. **b**, Quantification of volume change for unpaired (top; $n = 24/27$ suprathreshold (supra), $n = 25/29$ sub) (cells/spines) and paired (bottom; $n = 35/47$) (cells/spine pairs) stimuli. Black and grey triangles indicate suprathreshold and subthreshold stimuli, respectively. **c**, Effect of 0.25 μg ml^{-1} TrkB-Ig on synaptic crosstalk. $n = 6/13$ (cells/spine pairs). **d**, Effect of 0.125 μM 1NMPP1 on synaptic crosstalk in *Trkb*^{F616A} mice. $n = 5/10$ (cells/spine pairs). **e**, Effect of the Rac1 inhibitor NSC-23766 (NSC; 15 μM) on synaptic crosstalk. $n = 6/10$ (cells/spine pairs). **f**, Effect of 20 ng ml^{-1} BDNF on spine volume change after a subthreshold stimulus. $n = 6/6$ sub, $n = 6/7$ sub plus BDNF. **g**, Summary of **b**–**f**, with the addition of the Rac1 inhibitor EHT-1864 (EHT; $n = 7/12$; cells/spine pairs) and *Trkb*^{F616A} cells in the absence of 1NMPP1 (*Trkb*^{F616A} ctrl; $n = 5/7$; cells/spine pairs) showing averages of transient (1–2 min after stimulation) and sustained (>10 min after stimulation) phases of sLTP. Left, spines stimulated with a threshold stimulus. Right, spines stimulated with an unpaired or paired subthreshold stimulus. *$P < 0.05$ (Dunnet's test, versus the 'paired' subthreshold stimulus standard the crosstalk 'control'); #$P < 0.05$ (two-tailed t-test, versus each condition's paired crosstalk control).

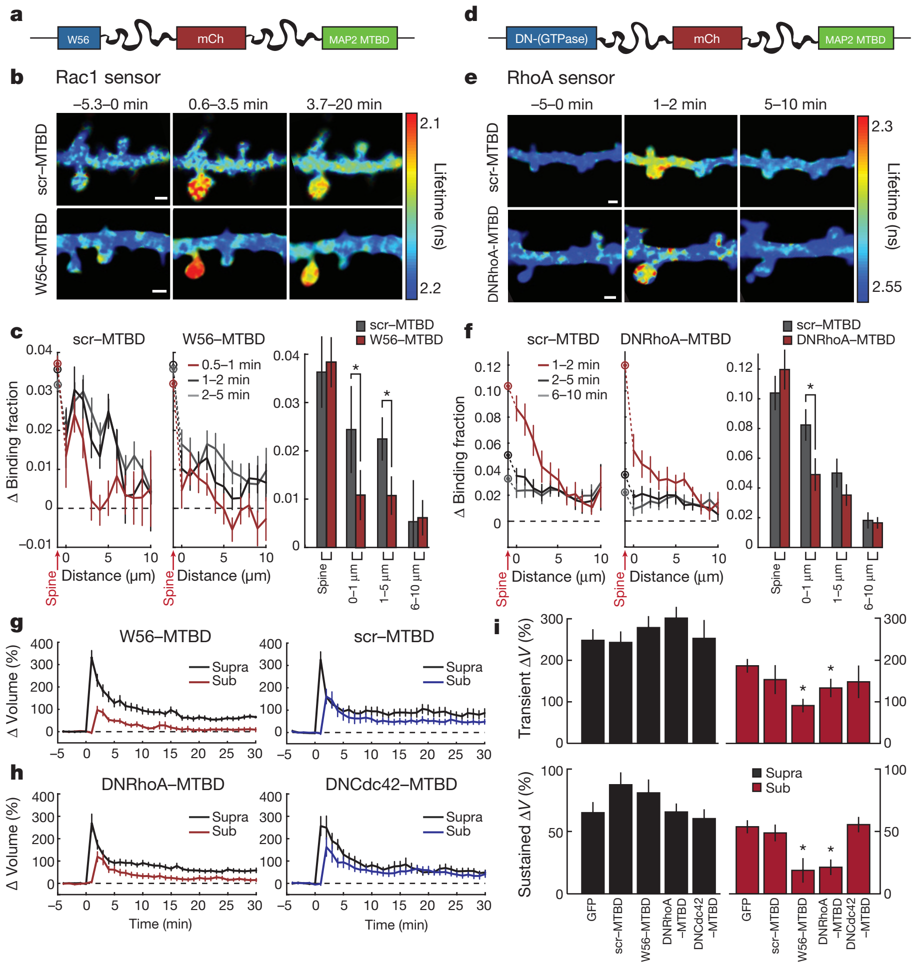


Figure 3 | Inhibition of signal spreading of Rac1 and RhoA prevents synaptic crosstalk. a, Schematic of the dendritic Rac1 inhibitor construct. W56 is fused to mCh and the microtubule binding domain (MTBD) of human MAP2 via intermediate linker sequences (wavy lines). **b**, Representative 2pFLIM images of the effect of W56–mCh–MTBD (W56–MTBD; bottom) or scrambled control (scr–MTBD; top) on the activation profile of the Rac1 sensor. Scale bars, 1 μm. **c**, Quantification of the spreading of Rac1 in scrambled control (left; $n = 10/17$) versus W56–mCh–MTBD (middle; $n = 22/36$) (cells/dendrite). Data represent the change in binding fraction in the dendrite as a function of distance from the stimulated spine (lines) and in the stimulated spines (circles) in the indicated time epochs. Right, Rac1 signal spreading in the presence of W56–mCh–MTBD across the indicated spatial windows at 1–2 min after stimulation. $^*P < 0.05$, two-tailed t-test. **d**, Schematic of the dendritic RhoA inhibitor, dominant-negative (DN) RhoA–mCh–MTBD. **e**, Representative 2pFLIM images of the effect of DNRhoA–mCh–MTBD expression on the activation profile of the RhoA sensor (bottom) versus scrambled control (top). Scale bars, 1 μm. **f**, Quantification of the spreading of RhoA in control ($n = 8/17$) versus DNRhoA–mCh–MTBD ($n = 11/18$) (cells/dendrite). Right, summary of RhoA spreading data. $^*P < 0.05$, two-tailed t-test. **g**, Effect of W56–mCh–MTBD (left; $n = 30/47$) versus scr–mCh–MTBD (right; $n = 14/17$) (cells/spine pairs). expression on synaptic crosstalk. **h**, Effect of DNRhoA–mCh–MTBD (left; $n = 14/26$) and DNCdc42–mCh–MTBD (right; $n = 10/15$) (cells/spine pairs) on synaptic crosstalk. **i**, Summary of crosstalk experiments. Right, averages for the crosstalk spine. Left, averages of the first ('LTP') spine. $^*P < 0.05$, Dunnet's test, compared to eGFP control.

in the stimulated spines (Extended Data Fig. 6d–g). Likewise, while a high concentration of the Rac1 inhibitor NSC-23766 significantly reduced sLTP (Extended Data Fig. 4a, c), a small concentration (15 μM) inhibited crosstalk without significantly affecting sLTP ($\Delta V_{\text{supra sustained}} = 48 \pm 12\%$, $\Delta V_{\text{sub sustained}} = 12 \pm 10\%$) (Fig. 2e, g). This NSC-23766 concentration also trended towards decreasing the Rac1 activation in the shaft (Extended Data Fig. 6h, i). These data suggest that BDNF initiates a signalling cascade capable of lowering the threshold of structural plasticity, consistent with previous reports implicating BDNF in similar phenomena[22]. Indeed, exogenous BDNF (20 ng ml^{-1}) application for ~10–15 min allowed a subthreshold stimulus alone to induce sLTP ($\Delta V_{\text{sub sustained}} = 119 \pm 31\%$) (Fig. 2f, g). Notably, BDNF application alone was sufficient to activate Rac1 and Cdc42, but had no significant effect on spine volume (Extended Data Fig. 7), suggesting that BDNF is facilitative of, but insufficient for, sLTP. Taken together, these results suggest that the BDNF–TrkB–Rac1 signalling initiated in a single spine during sLTP induction facilitates sLTP in nearby spines, allowing synaptic crosstalk.

To probe the requirement of heterosynaptic Rac1 activity for synaptic crosstalk more specifically, we devised a strategy to interrupt Rac1 activity spreading out of the stimulated spine during sLTP: we restricted the Rac1 inhibitory peptide W56 (ref. 23) to dendrites by fusing it to the microtubule binding domain (MTBD) of MAP2, concentrating the inhibitor to dendritically enriched microtubules[24,25] (Fig. 3a, Extended Data Fig. 8a). This construct, W56–mCh–MTBD (in which mCh denotes mCherry), was concentrated in dendritic shafts and excluded from spines (Extended Data Fig. 8a). We found that this inhibitor significantly reduced Rac1 activation in the dendritic shaft, while largely preserving its activation in the stimulated spine (Fig. 3b, c). By contrast, W56–mCh without MTBD resulted in a more global reduction in Rac1 activation (Extended Data Fig. 8c, d). When the W56 peptide was scrambled (scr–mCh–MTBD), Rac1 activity spreading was normal (Fig. 3b, c). Finally, W56–mCh–MTBD did not change the spatial spreading of RhoA (Extended Data Fig. 8e). These results suggest that W56–mCh–MTBD specifically inhibits the spreading of Rac1 activity into dendrites.

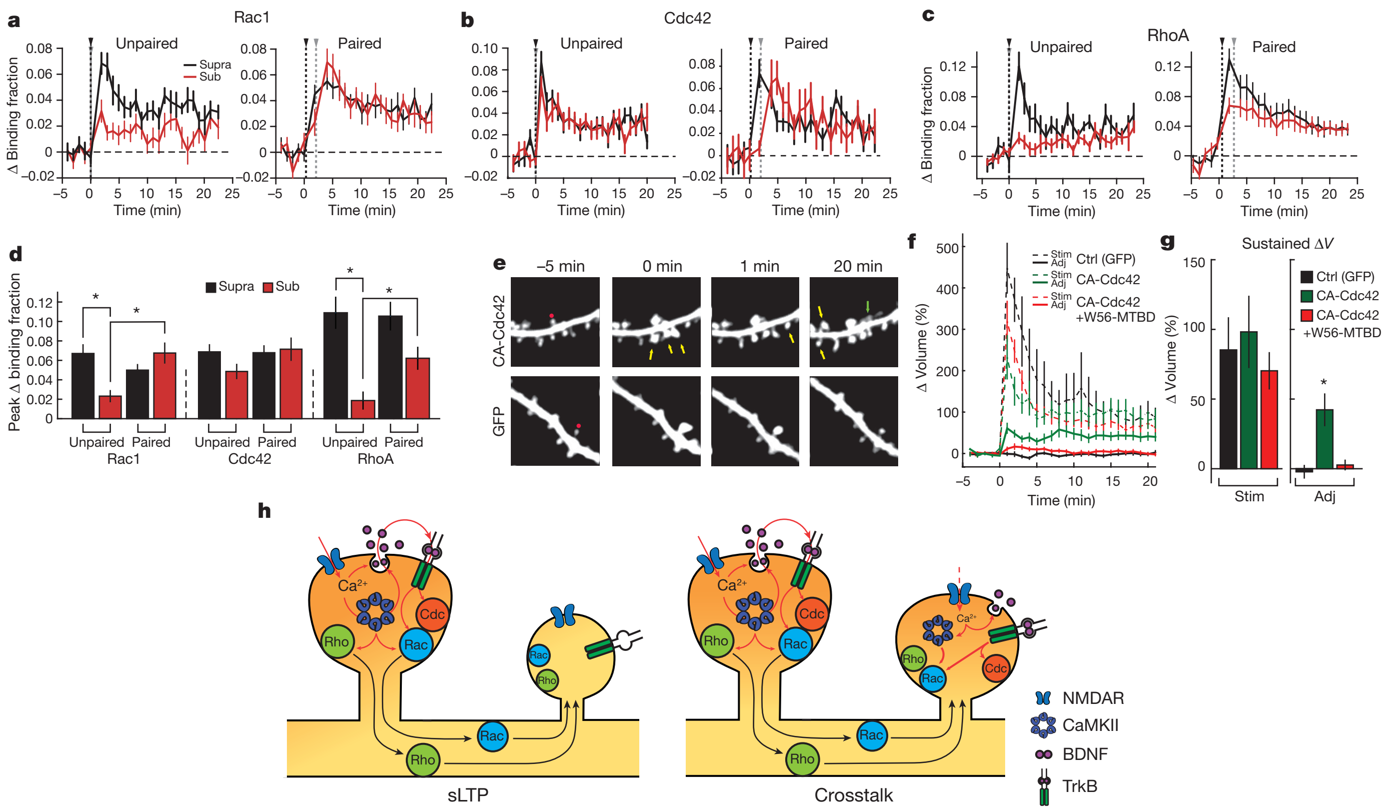


Figure 4 | Signal spreading provides additive activation of high-threshold signals during synaptic crosstalk. a, Comparison of Rac1 activation in response to unpaired (left; suprathreshold (black) $n = 12/12$, subthreshold (red) $n = 15/15$; cells/spines) and paired crosstalk (right; $n = 5/9$; cells/spine pairs) stimuli. Black and grey arrows indicate supra/subthreshold stimulus onset. **b**, Same as **a** for Cdc42 activity (left: unpaired; suprathreshold $n = 14/18$, subthreshold $n = 14/20$; right: paired; $n = 9/10$). **c**, Same as **a** and **b** for RhoA activity (left: unpaired; suprathreshold $n = 9/9$, subthreshold $n = 17/22$; right: paired; $n = 13/24$). **d**, Summary of data from **a**–**c**. Bars represent average peak activity. For unpaired conditions, the peak is the average of the first 2 min following uncaging. For paired conditions, the peak for suprathreshold spines corresponds to the first 2 min after the stimulus, while the peak of the crosstalk spines was based on the timing of maximal response for each GTPase (Rac1 and Cdc42: 1–2 min after the second stimulus; RhoA: 1–2 min after the first stimulus) to account for signal spreading. $*P < 0.05$, analysis of variance (ANOVA) with Tukey–Kramer's post-hoc test. **e**, Top, representative two-photon images of the effect of expressing CA-Cdc42 on the spine-specificity of sLTP. Yellow arrows mark non-targeted spines that increased in intensity after uncaging at the target (red circle) spine. The green arrow represents a new spine. Bottom, representative images of a typical sLTP experiment in which there is no heterosynaptic effect of uncaging. Image size, $12 \times 12\,\mu$m. **f**, Quantification of the average change in nearby ($<\sim 5\,\mu$m) spine volume after glutamate uncaging in control (black curve; $n = 13$ cells/13 targeted spines/112 neighbouring spines), CA-Cdc42-expressing (green curve; $n = 9$ cells/14 targeted spines/102 neighbouring spines), and CA-Cdc42-plus W56–mCh–MTBD-expressing (red curve; $n = 5$ cells/11 targeted spines/78 neighbouring spines) cells. **g**, Summary of data in **f**. $*P < 0.05$, ANOVA with Tukey–Kramer's test. **h**, Proposed model. Calcium through NDMARs from a strong stimulus activates CaMKII, supporting autocrine BDNF–TrkB signalling, and conferring actin cytoskeleton remodelling through Rac1 and Cdc42. In parallel, CaMKII-dependent RhoA activation also acts on the actin cytoskeleton. The combination of the three GTPases is necessary to produce sLTP. Rac1 and RhoA activity spread out of the stimulated spine into the dendrite and surrounding spines, which is insufficient to produce sLTP. However, even a weak stimulus (red dotted line), can cause sufficient BDNF release to activate Cdc42, complementing the activity of Rac1 and RhoA and allowing synaptic crosstalk.

Consistent with the hypothesis that Rac1 activity spreading is required for synaptic crosstalk, the expression of W56–mCh–MTBD reduced crosstalk without significantly affecting sLTP (Fig. 3g). By contrast, a similar expression level of scr–mCh–MTBD affected neither sLTP nor crosstalk (Fig. 3g, Extended Data Fig. 8b). To test the specificity of this manipulation further, we tethered the Rac1-specific GTPase-activating protein ARHGAP15 (ref. 26) to MTBD. This construct also reduced Rac1 activity spreading without significantly affecting activation in the spine, and inhibited crosstalk without affecting sLTP (Extended Data Fig. 8f–k). Thus, activation of Rac1 in the dendritic shaft is necessary for synaptic crosstalk.

Since RhoA activation also spreads into nearby spines during sLTP, we next tested whether this BDNF-independent signal is also required for synaptic crosstalk. To do this, we tethered dominant-negative RhoA (DNRhoA) to mCh–MTBD (Fig. 3d). Expressing DNRhoA–mCh–MTBD significantly reduced a portion of RhoA activity spreading (Fig. 3e, f), and also blocked crosstalk without significantly affecting sLTP ($\Delta V_{\text{supra sustained}} = 66 \pm 6\%$, $\Delta V_{\text{sub sustained}} = 21 \pm 6\%$) (Fig. 3h, i). Importantly, using this same strategy against Cdc42 (the activity of which is compartmentalized in spines) had no effect on either sLTP or crosstalk ($\Delta V_{\text{supra sustained}} = 60 \pm 7\%$, $\Delta V_{\text{sub sustained}} = 56 \pm 6\%$) (Fig. 3h, i), suggesting that this approach is only effective when targeting proteins with diffusive activation profiles. Taken together, these data suggest that the convergence of BDNF-dependent Rac1 signalling and BDNF-independent RhoA signalling at nearby spines primes these regions for facilitated structural plasticity.

To examine how crosstalk is achieved despite signal spreading conferring activation of only a subset of GTPases, we measured the activation of Rho GTPases in response to both unpaired and paired suprathreshold and subthreshold stimuli (Fig. 4a–d; see Extended Data Fig. 9a, b for response variability and Extended Data Figs 9c and 10 for volume curves of subthreshold stimuli). We found that both RhoA and Rac1, which show diffusive activity profiles, were only weakly activated by unpaired subthreshold stimuli (Fig. 4a, c). However, when paired with nearby suprathreshold stimuli, signal spreading supplied the crosstalk spines with additional Rac1 and RhoA activation (Fig. 4a, c, d). By contrast, Cdc42 was strongly activated by a subthreshold stimulus in both the unpaired and paired conditions, thus explaining how

nearby spines achieve the required levels of Cdc42 activation during crosstalk (Fig. 4b, d). Thus, through a combination of low-threshold, spine-specific Cdc42 activation and signal spreading of RhoA/Rac1 provided by nearby sLTP, Rho GTPase signalling complementation facilitates detection of weak synaptic activity proximal to sites of sLTP. Notably, a weak stimulus also caused a modest level of BDNF release (as measured by BDNF fused to the pH-sensitive fluorophore superecliptic pHluorin; BDNF–SEP) and TrkB activation, supporting the notion that this pathway is still functional during a weak stimulus (Extended Data Fig. 9d, e).

The compartmentalized nature of Cdc42 activity in spines probably serves to prevent nonspecific structural plasticity at nearby inactive synapses. Consistent with this model, stimulation of a single spine on cells expressing constitutively active (CA)-Cdc42 caused significant enlargement of surrounding, unstimulated spines ($\Delta V_{\text{nearby sustained}} = 42 \pm 11\%$) (Fig. 4e–g). This is in sharp contrast to control conditions, in which nearby spines show no average change in volume ($\Delta V_{\text{nearby sustained}} = -2 \pm 4\%$) (Fig. 4e–g). Importantly, sLTP in the stimulated spines was similar to that of control (Fig. 4e–g). Furthermore, when CA-Cdc42 was co-expressed with W56–mCh–MTBD, sLTP in nearby spines was suppressed without affecting sLTP at the stimulated spines ($\Delta V_{\text{nearby sustained}} = 3 \pm 3\%$; $\Delta V_{\text{stim sustained}} = 70 \pm 13\%$) (Fig. 4e–g), suggesting that the heterosynaptic effects imparted by CA-Cdc42 depend on Rac1 signal spreading. Thus, removing the compartmentalization of Cdc42 activation degrades the input-specificity of sLTP, likely by complementing the typical priming effects of diffusive Rac1 and RhoA signals.

Collectively, our data suggest that simultaneous activation of Rac1, Cdc42 and RhoA predicts the occurrence of sLTP, while activation of a subset of these proteins primes spines for sLTP (Fig. 4h). The activation of both Cdc42 and Rac1 required postsynaptic, autocrine BDNF, and conveyed the occurrence of sLTP over both spine-specific (Cdc42) and heterosynaptic (Rac1) domains. The spreading of BDNF–TrkB-mediated Rac1 signalling out of the stimulated spine was necessary for facilitating sLTP in nearby spines, consistent with the known pro-plasticity properties of BDNF[22,27–29]. The spreading of BDNF-independent RhoA activation was also necessary for synaptic crosstalk, suggesting that both BDNF-dependent and -independent signalling pathways must be coordinated to achieve this phenomenon. The combination of spine-specific, low-threshold Cdc42 activation together with diffusive, high-threshold RhoA and Rac1 activation is perfectly positioned to simultaneously achieve both spine-specific homosynaptic sLTP and facilitation of sLTP in surrounding spines. Thus, our model, based on the coincident activation of three small GTPase molecules Rac, Rho and Cdc42, together with the autocrine BDNF signalling, provides a unified theory for both homo- and heterosynaptic plasticity.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 3 June; accepted 22 August 2016.
Published online 28 September 2016.

1. Tashiro, A., Minden, A. & Yuste, R. Regulation of dendritic spine morphology by the rho family of small GTPases: antagonistic roles of Rac and Rho. *Cereb. Cortex* **10,** 927–938 (2000).
2. Dietz, D. M. *et al.* Rac1 is essential in cocaine-induced structural plasticity of nucleus accumbens neurons. *Nat. Neurosci.* **15,** 891–896 (2012).
3. Murakoshi, H., Wang, H. & Yasuda, R. Local, persistent activation of Rho GTPases during plasticity of single dendritic spines. *Nature* **472,** 100–104 (2011).
4. Okamoto, K., Nagai, T., Miyawaki, A. & Hayashi, Y. Rapid and persistent modulation of actin dynamics regulates postsynaptic reorganization underlying bidirectional plasticity. *Nat. Neurosci.* **7,** 1104–1112 (2004).
5. Nakayama, A. Y., Harms, M. B. & Luo, L. Small GTPases Rac and Rho in the maintenance of dendritic spines and branches in hippocampal pyramidal neurons. *J. Neurosci.* **20,** 5329–5338 (2000).
6. Haditsch, U. *et al.* A central role for the small GTPase Rac1 in hippocampal plasticity and spatial learning and memory. *Mol. Cell. Neurosci.* **41,** 409–419 (2009).
7. Lai, K.-O. O. *et al.* TrkB phosphorylation by Cdk5 is required for activity-dependent structural plasticity and spatial memory. *Nat. Neurosci.* **15,** 1506–1515 (2012).
8. Il Hwan, K., Hong, W., Scott, H. S. & Ryohei, Y. Loss of Cdc42 leads to defects in synaptic plasticity and remote memory recall. *eLife* **3,** e02839 (2014).
9. Machacek, M. *et al.* Coordination of Rho GTPase activities during cell protrusion. *Nature* **461,** 99–103 (2009).
10. McAllister, A. K. Biolistic transfection of neurons. *Sci. STKE* **2000,** pl1 (2000).
11. Stoppini, L., Buchs, P. A. & Muller, D. A simple method for organotypic cultures of nervous tissue. *J. Neurosci. Methods* **37,** 173–182 (1991).
12. Yasuda, R. *et al.* Supersensitive Ras activation in dendrites and spines revealed by two-photon fluorescence lifetime imaging. *Nat. Neurosci.* **9,** 283–291 (2006).
13. Matsuzaki, M., Honkura, N., Ellis-Davies, G. C. & Kasai, H. Structural basis of long-term potentiation in single dendritic spines. *Nature* **429,** 761–766 (2004).
14. Honkura, N., Matsuzaki, M., Noguchi, J., Ellis-Davies, G. C. & Kasai, H. The subspine organization of actin fibers regulates the structure and plasticity of dendritic spines. *Neuron* **57,** 719–729 (2008).
15. Harward, S. C. *et al.* Autocrine BDNF–TrkB signalling within a single dendritic spine. *Nature* http://dx.doi.org/10.1038/nature19766 (2016).
16. Lu, W. *et al.* Subunit composition of synaptic AMPA receptors revealed by a single-cell genetic approach. *Neuron* **62,** 254–268 (2009).
17. Luikart, B. W. *et al.* TrkB has a cell-autonomous role in the establishment of hippocampal Schaffer collateral synapses. *J. Neurosci.* **25,** 3774–3786 (2005).
18. Miyamoto, Y., Yamauchi, J., Tanoue, A., Wu, C. & Mobley, W. C. TrkB binds and tyrosine-phosphorylates Tiam1, leading to activation of Rac1 and induction of changes in cellular morphology. *Proc. Natl Acad. Sci. USA* **103,** 10444–10449 (2006).
19. Harvey, C. D. & Svoboda, K. Locally dynamic synaptic learning rules in pyramidal neuron dendrites. *Nature* **450,** 1195–1200 (2007).
20. Harvey, C. D., Yasuda, R., Zhong, H. & Svoboda, K. The spread of Ras activity triggered by activation of a single dendritic spine. *Science* **321,** 136–140 (2008).
21. Chen, X. *et al.* A chemical-genetic approach to studying neurotrophin signaling. *Neuron* **46,** 13–21 (2005).
22. Huber, K. M., Sawtell, N. B. & Bear, M. F. Brain-derived neurotrophic factor alters the synaptic modification threshold in visual cortex. *Neuropharmacology* **37,** 571–579 (1998).
23. Gao, Y., Xing, J., Streuli, M., Leto, T. L. & Zheng, Y. Trp(56) of rac1 specifies interaction with a subset of guanine nucleotide exchange factors. *J. Biol. Chem.* **276,** 47530–47541 (2001).
24. Korobova, F. & Svitkina, T. Molecular architecture of synaptic actin cytoskeleton in hippocampal neurons reveals a mechanism of dendritic spine morphogenesis. *Mol. Biol. Cell* **21,** 165–176 (2010).
25. Zhong, H. *et al.* Subcellular dynamics of type II PKA in neurons. *Neuron* **62,** 363–374 (2009).
26. Seoh, M. L., Ng, C. H., Yong, J., Lim, L. & Leung, T. ArhGAP15, a novel human RacGAP protein with GTPase binding property. *FEBS Lett.* **539,** 131–137 (2003).
27. Tanaka, J. *et al.* Protein synthesis and neurotrophin-dependent structural plasticity of single dendritic spines. *Science* **319,** 1683–1687 (2008).
28. Korte, M. *et al.* Hippocampal long-term potentiation is impaired in mice lacking brain-derived neurotrophic factor. *Proc. Natl Acad. Sci. USA* **92,** 8856–8860 (1995).
29. Bosch, M. *et al.* Structural and molecular remodeling of dendritic spine substructures during long-term potentiation. *Neuron* **82,** 444–459 (2014).
36. He, X.P. *et al.* Conditional deletion of TrkB but not BDNF prevents epileptogenesis in the kindling model. *Neuron* **43,** 31–42 (2004).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was supported by grants from NIMH (R01MH080047 (R.Y.), R01NS068410 (R.Y.)), NINDS (F31NS078847 (S.C.H.), R01NS05621 (J.O.M.), DP1NS096787 (R.Y.)), JSPS KAKENHI (H.M.), JST PRESTO (H.M.), the Wakeman Fellowship at Duke University (S.C.H.) and Max Planck Florida Institute (R.Y.). We thank A. E. West, S. Soderling, and F. Wang for guidance and valuable discussion, and Yasuda and McNamara laboratory members for discussion. We also thank R. Puranam for isolating MTBD. Finally, we thank D. Kloetzer for managing the laboratory.

Author Contributions N.G.H., S.C.H., J.O.M. and R.Y. designed the experiments. N.G.H., S.C.H. and H.M. developed the sensor and inhibitors; N.G.H. and S.C.H. collected the data with assistance from C.E.H.; N.G.H., S.C.H., C.E.H. and R.Y. analysed the data; N.G.H., S.C.H., J.M.O. and R.Y. wrote the paper. All authors discussed the results and comments on this manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.Y. (ryohei.yasuda@mpfi.org).

Reviewer Information *Nature* thanks B. Bingol, H. Zhang and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Reagents. Human recombinant BDNF was purchased from Millipore; D-2-amino-5-phosphonovalerate (D-AP5) and NSC-23766 were from Tocris; 1'-naphthylmethyl-4-amino-1-*tert*-butyl-3-(*p*-methylphenyl)pyrazolo[3,4-*d*] pyrimidine (1NMPP1) was from Santa Cruz and Shanghai Institute of Materia Medica, Chinese Academy of Sciences; and TrkB-Ig was a gift from Regeneron. The tat-CN21 peptide (YGRKKRRQRRRKRPPKLGQIGRSKRVVIEDDR) was synthesized by GenScript.

Animals. All animal procedures were approved by the Duke Univeristy School of Medicine Animal Care and Use Committee. Both male and female rats and mice were used. *Trkb*^{F616A} mutant mice were provided by D. Ginty[21]. *Bdnf*$^{fl/fl}$ and *Trkb*$^{fl/fl}$ were provided by L. Parada[17,36]. *Rac1*$^{fl/fl}$ animals were acquired from C. Brakebusch[30]. The genotype of each animal was verified before and after preparing slices using PCR of genomic DNA isolated from tail DNA before and slice samples after.

Plasmids. Plasmids containing human RAC1 and PAK1(65–118) are gifts from M. Matsuda and S. Soderling, respectively. The Pak GTPase binding domain of PAK2 (PBD2) was prepared by introducing mutations L77P and S115L into PAK1(60–118) using a Site-Directed Mutagenesis kit (Stratagene). W56–mCh–MTBD was prepared by amplifying the Rac1 inhibitory peptide W56 (ref. 23) using overhang PCR with a C-terminal linker (GGGGGGGGGGGGGGGGGGGGGGGMADQLTEEWHRGTAGPGS) and inserting it into pCAG-mCh-mCh (ref. 3) by removing the first mCh with EcoRI and KpnI restriction digest and replacing it with the W56-linker amplicon, creating pCAG-W56-(linker)-mCh. In parallel, the MTBD of human MAP2 (272-end)[25] was isolated from a human cDNA library and PCR amplification. This amplicon was then further amplified with overhang PCR to contain a linker (same as above), and then inserted into pCAG-mCh-mCh using BamHI-NotI restriction digest to produce pCAG-mCh-(Linker)-MTBD. The two constructs were then combined using BamHI plus NotI restriction digest to create pCAG-W56-(linker)-mCh-(linker)-MTBD. The scrambled variant of W56–MTBD was created by randomly re-ordering the residues of W56. ARHGAP15-mCh-MTBD was made by inserting ARHGAP15 (1–723; Addgene plasmid 38903) into the -mCh-MTBD sequence described above by adding EcoRI and KpnI sites at the N and C terminus, respectively.

DNRhoA and DNCdc42 variants of the X-mCh-MTBD construct were prepared by first incorporating an MfeI digestion site on the 3′ end of W56, then removing W56 by digestion with NheI/MfeI and insertion of the dominant-negative construct.

Preparation. Hippocampal slices were prepared from postnatal day 5–7 rats or mice in accordance with the animal care and use guidelines of Duke University Medical Centre. In brief, we deeply anaesthetized the animal with isoflurane, after which the animal was quickly decapitated and the brain removed. The hippocampi were isolated and cut into 350-μm sections using a McIlwain tissue chopper. Hippocampal slices were plated on tissue culture inserts (Millicell) fed by tissue medium (for 2.5 l: 20.95 g MEM, 17.9 g HEPES, 1.1 g $NaHCO_3$, 5.8 g D-glucose, 120 μl 25% ascorbic acid, 12.5 ml L-glutamine, 2.5 ml insulin, 500 ml horse serum, 5 ml 1 M $MgSO_4$, 2.5 ml 1 M $CaCl_2$). Slices were incubated at 35 °C in 3% CO_2.

After 1–2 weeks in culture, CA1 pyramidal neurons were transfected with ballistic gene transfer using gold beads (8–12 mg) coated with plasmids containing 30 μg of total cDNA (Rac1 sensor, donor:acceptor = 1:2; eGFP + W56–MTBD, 5:1; Rac1 sensor + W56–MTBD, donor:acceptor:inhibitor = 2:4:1; TrkB sensor, donor:acceptor = 1:1; Cdc42 sensor, donor:acceptor = 1:1; RhoA sensor, donor:acceptor = 1:1). Cells expressing only eGFP were imaged 1–5 days after transfection, cells expressing TrkB were imaged 1–2 days after transfection, and all other plasmid combinations were imaged 2–5 days after transfection.

For structural plasticity experiments, conditional knockout slices (*Bdnf*$^{fl/fl}$ and *Rac1*$^{fl/fl}$) were transfected with either eGFP alone or eGFP and tdTomato-Cre (1:1) for 3–7 days before imaging. For sensor experiments in these slices, the sensors were used in the ratios listed above with an amount of Cre recombinase equal to the amount of donor DNA. The presence of Cre was confirmed by nuclear-localized tdTomato signal.

HEK293T cells (ATCC) were cultured in DMEM supplemented with 10% fetal calf serum at 37 °C in 5% CO_2. Transfection was performed at ~50–90% cell confluency using Lipofectamine (Invitrogen) and 2 μg ml^{-1} of total cDNA/35 mm dish, following the ratios listed above. Cells were used as an expression platform only, and were thus not rigorously tested for potential contamination from other cell lines.

2pFLIM. FRET imaging using a custom-built two-photon fluorescence lifetime imaging microscope was performed as previously described[3,31,32]. Two-photon imaging was performed using a Ti-sapphire laser (MaiTai, Spectraphysics) tuned to a wavelength of 920 nm, allowing simultaneous excitation of eGFP and mCh. All samples were imaged using <2 mW laser power measured at the objective. Fluorescence emission was collected using an immersion objective (60×, numerical aperture 0.9, Olympus), divided with a dichroic mirror (565 nm), and detected with two separate photoelectron multiplier tubes (PMTs) placed downstream of two wavelength filters (Chroma, HQ510-2p to select for green and HQ620/90-2p to select for red). The green channel was fitted with a PMT having a low transfer time spread (H7422-40p; Hamamatsu) to allow for fluorescence lifetime imaging, while the red channel was fitted with a wide-aperture PMT (R3896; Hamamatsu). Photon counting for fluorescence lifetime imaging was performed using a time-correlated single photon counting board (SPC-150; Becker and Hickl) controlled with custom software[31], while the red channel signal was acquired using a separate data acquisition board (PCI-6110) controlled with Scanimage software[33].

Two-photon glutamate uncaging. A second Ti-sapphire laser tuned at a wavelength of 720 nm was used to uncage 4-methoxy-7-nitroindolinyl-caged-L-glutamate (MNI-caged glutamate) in extracellular solution with a train of 4–6 ms, 4–5 mW pulses (30 times at 0.5 Hz) near a spine of interest ('sLTP stimulus'). Experiments were performed in Mg^{2+} fee artificial cerebral spinal fluid (ACSF; 127 mM NaCl, 2.5 mM KCl, 4 mM $CaCl_2$, 25 mM $NaHCO_3$, 1.25 mM NaH_2PO_4 and 25 mM glucose) containing 1 μM tetrodotoxin (TTX) and 4 mM MNI-caged L-glutamate aerated with 95% O_2 and 5% CO_2 at 30 °C, as described previously. Subthreshold stimuli were delivered using a train of 1 ms, 4–5 mW pulses (30 times at 0.5 Hz). Crosstalk experiments were performed by first delivering an sLTP stimulus (4–6 ms), then delivering a subthreshold stimulus to a nearby (~2–5 μm) spine on the same dendrite ~90 s later, as previously described[19]. Anywhere from 1–5 spines were stimulated per cell, and a maximum of 3 crosstalk experiments were performed on a single cell.

Spine volume analysis. Spine volume was calculated as the background-subtracted integrated fluorescence intensity over a region of interest around the dendritic spine head (fluorescence, F). Change in spine volume was measured as F/F_0, in which F_0 is the average fluorescence intensity before stimulation. Analysis of two-photon images outside of the context of 2pFLIM was performed in ImageJ.

All experiments involving dendritic inhibitor constructs were performed in a blinded fashion until the experiments were complete (when the groups significantly diverged, or until 15–20 individual experiments across at least three cells were complete, whichever came first).

2pFLIM data analyses. To measure the fraction of donor bound to acceptor, we fit a fluorescence lifetime curve summing all pixels over a whole image with a double exponential function convolved with the Gaussian pulse response function:

$$F(t) = F_0[P_D H(t, t_0, \tau_D, \tau_G) + P_{AD} H(t, t_0, \tau_{AD}, \tau_G)] \quad (1)$$

where τ_{AD} is the fluorescence lifetime of donor bound with acceptor, P_D and P_{AD} are the fraction of free donor and donor bound with acceptor, respectively, and $H(t)$ is a fluorescence lifetime curve with a single exponential function convolved with the Gaussian pulse response function:

$$H(t, t_0, t_D, t_G) = \frac{1}{2}\exp\left(\frac{\tau_G^2}{2\tau_D^2} - \frac{t - t_0}{\tau_D}\right)\operatorname{erfc}\left(\frac{\tau_G^2 - \tau_D(t - t_0)}{2\tau_D\tau_G}\right)$$

in which τ_D is the fluorescence lifetime of the free donor, τ_G is the width of the Guassian pulse response function, F_0 is the peak fluorescence before convolution and t_0 is the time offset, and erfc is the error function.

We fixed τ_D to the fluorescence lifetime obtained from free eGFP (2.6 ns). To generate the fluorescence lifetime image, we calculated the mean photon arrival time, $\langle t\rangle$, in each pixel as:

$$\langle t\rangle = \int tF(t)dt \Big/ \int F(t)dt,$$

then, the mean photon arrival time is related to the mean fluorescence lifetime, $\langle\tau\rangle$, by an offset arrival time, t_0, which is obtained by fitting the whole image:

$$\langle\tau\rangle = \langle t\rangle - t_0$$

For small regions-of-interest (ROIs) in an image (spines or dendrites), we calculated the binding fraction (P_{AD}) as:

$$P_{AD} = \tau_D(\tau_D - \langle\tau\rangle)(\tau_D - \tau_{AD})^{-1}(\tau_D + \tau_{AD} - \langle\tau\rangle)^{-1}$$

Measurements of the affinity between Rho GTPases and RBDs. Polyhistidine-tagged super-folder GFP (sfGFP)–Rac1, mCh-PBD2 and their mutants were cloned into the pRSET bacterial expression vector (Invitrogen). Proteins were overexpressed in *Escherichia coli* (DH5α), purified with a Ni^+-nitrilotriacetate (NTA) column (HiTrap, GE Healthcare), and desalted with a desalting column (PD10, GE Healthcare) equilibrated with PBS. The concentration of the purified protein was measured by the absorbance of the fluorophore (sfGFP, $A_{489\,nm}$ = 83,000 cm^{-1} M^{-1} (ref. 34); mCh, $A_{587\,nm}$ = 72,000 cm^{-1} M^{-1} (ref. 35)).

Purified sfGFP–Rac1 was loaded with GppNHp (2',3'-*O*-*N*-methyl anthraniloyl-GppNHp) and GDP by incubating in the presence of tenfold molar excess of GppNHp and GDP in $MgCl_2$-free PBS containing 1 mM EDTA for 10 min, respectively. The reaction was terminated by adding 10 mM $MgCl_2$. sfGFP–Rac1 and mCh–PBD2 were mixed and incubated at room temperature for 20 min. FRET between sfGFP and mCh was measured under 2pFLIM, and the fraction of sfGFP–Rac1 bound to mCh–PDB2 was calculated by fitting the fluorescence lifetime curve with a double exponential function (equation (1)). The dissociation constant was obtained by fitting the relationship between the binding fraction and the concentration of mCh–PDB2 ([mCh–PDB2]) with a Michaelis–Menten function.

Statistical methods. Sample sizes for all experiments were chosen based on signal-to-noise ratios identified in pilot experiments.

The variances of all data were estimated and compared using Bartlett's test or Levene's test before further statistical analysis.

The distribution patterns of Rho GTPase sensor activity was determined by performing a Shapiro–Wilk test for normality on the peak response (the same points used for statistical comparisons). All of the sensors tested adhered to the null hypothesis, and thus are considered normally distributed. As such, parametric statistics were used to compare values of Rho GTPases response.

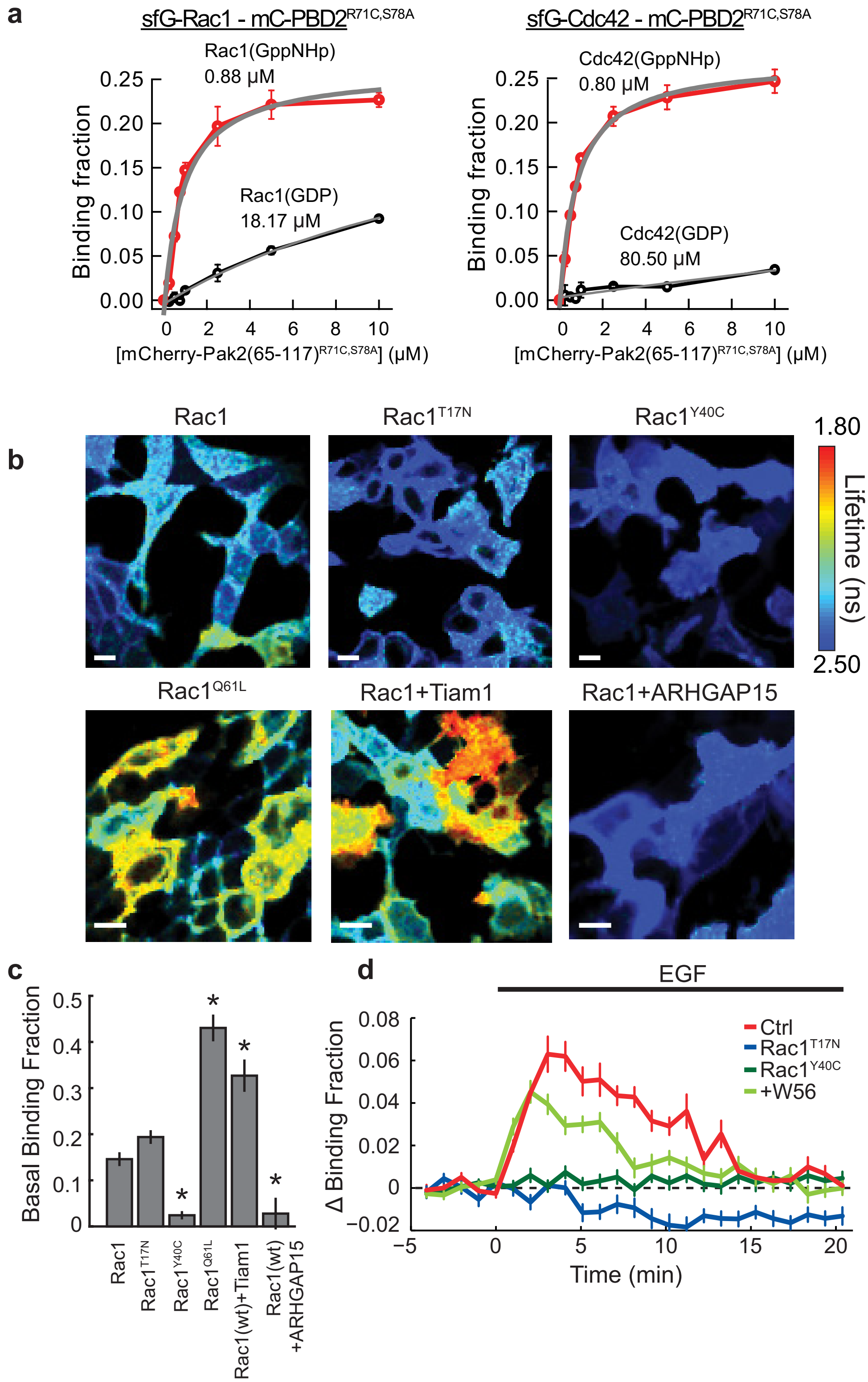
For multiple comparisons of sensor activity, data were first subjected to ANOVA, followed by a post-hoc test to determine statistical significance, according to the structure of the comparison being made. In cases where each condition is being compared to all other conditions in the group, the Tukey–Kramer method was used. In cases where each condition is being compared to a single control, Dunnet's test was used instead.

To compare values of non-normally distribution changes in spine volume, data were log-transformed to resolve skewness, then subjected to normal parametric statistics, as indicated in the figure legends. To support these statistical claims, non-parametric statistics were also applied to the original, non-transformed data using a Wilcoxon rank-sum test in place of *t*-tests, and the Kruskal–Wallis procedure in place of ANOVA, followed by a post-hoc analysis using Dunn's test. All of the data tested were significant by both of these approaches.

Data were only excluded if obvious signs of poor cellular health (for example, dendritic blebbing, spine collapse) were apparent.

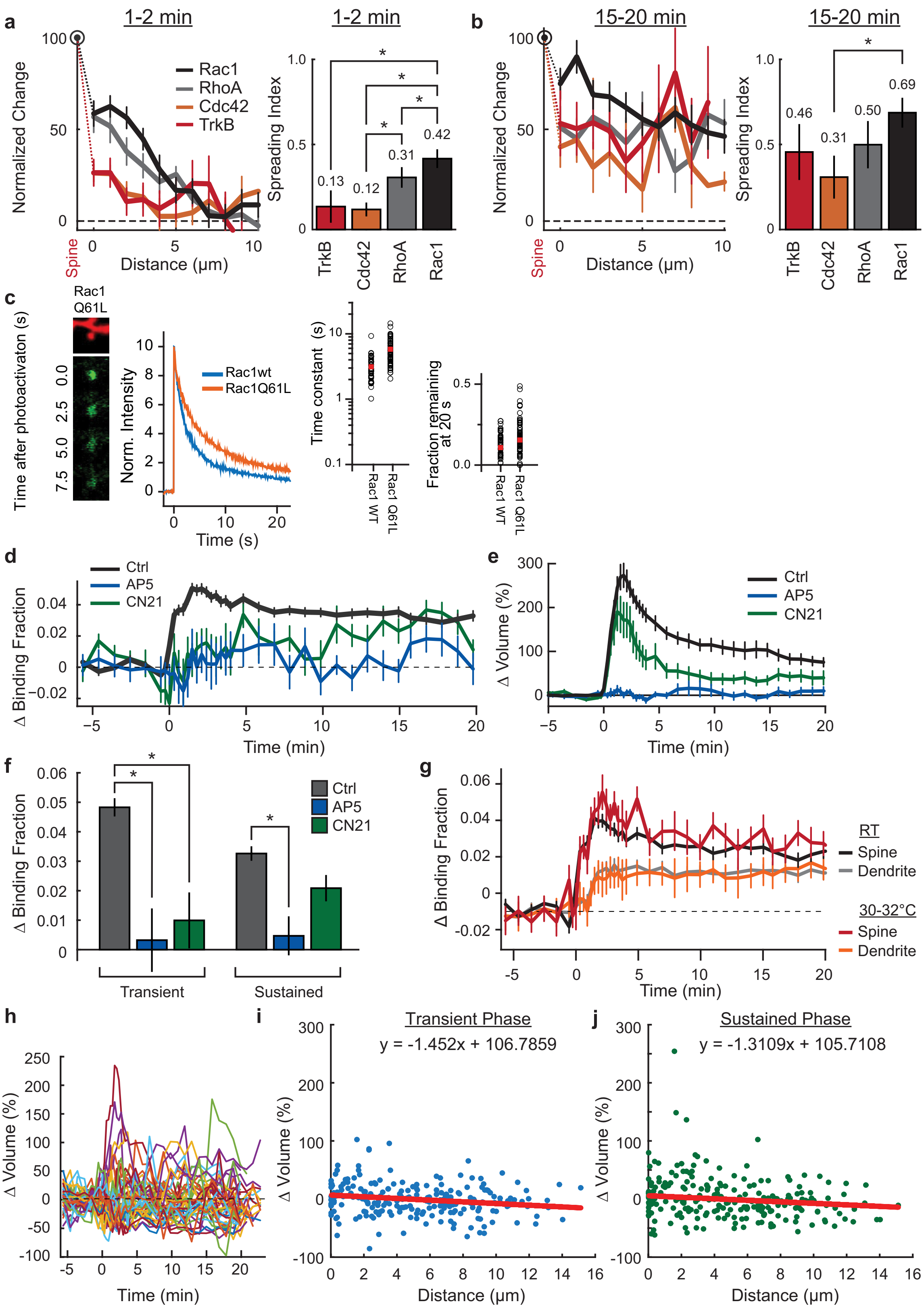
Crosstalk experiments comparing different genetic perturbations were performed in a blinded fashion. Experimenters were unblinded when either statistical significance was reached, or when experimental number was comparable to similar experiments that had reached statistical significance.

30. Chrostek, A. *et al.* Rac1 is crucial for hair follicle integrity but is not essential for maintenance of the epidermis. *Mol. Cell. Biol.* **26,** 6957–6970 (2006).
31. Yasuda, R. Imaging spatiotemporal dynamics of neuronal signaling using fluorescence resonance energy transfer and fluorescence lifetime imaging microscopy. *Curr. Opin. Neurobiol.* **16,** 551–561 (2006).
32. Murakoshi, H., Lee, S.-J. & Yasuda, R. Highly sensitive and quantitative FRET-FLIM imaging in single dendritic spines using improved non-radiative YFP. *Brain Cell Biol.* **36,** 31–42 (2008).
33. Pologruto, T. A., Sabatini, B. L. & Svoboda, K. ScanImage: flexible software for operating laser scanning microscopes. *Biomed. Eng. Online* **2,** 13 (2003).
34. Shaner, N. C. *et al.* Improved monomeric red, orange and yellow fluorescent proteins derived from *Discosoma* sp. red fluorescent protein. *Nat. Biotech.* **22,** 1567–1572 (2004).
35. Pédelacq, J. D., Cabantous, S., Tran, T. & Terwilliger, T. C. Engineering and characterization of a superfolder green fluorescent protein. *Nat. Biotech.* **24,** 79–88 (2005).



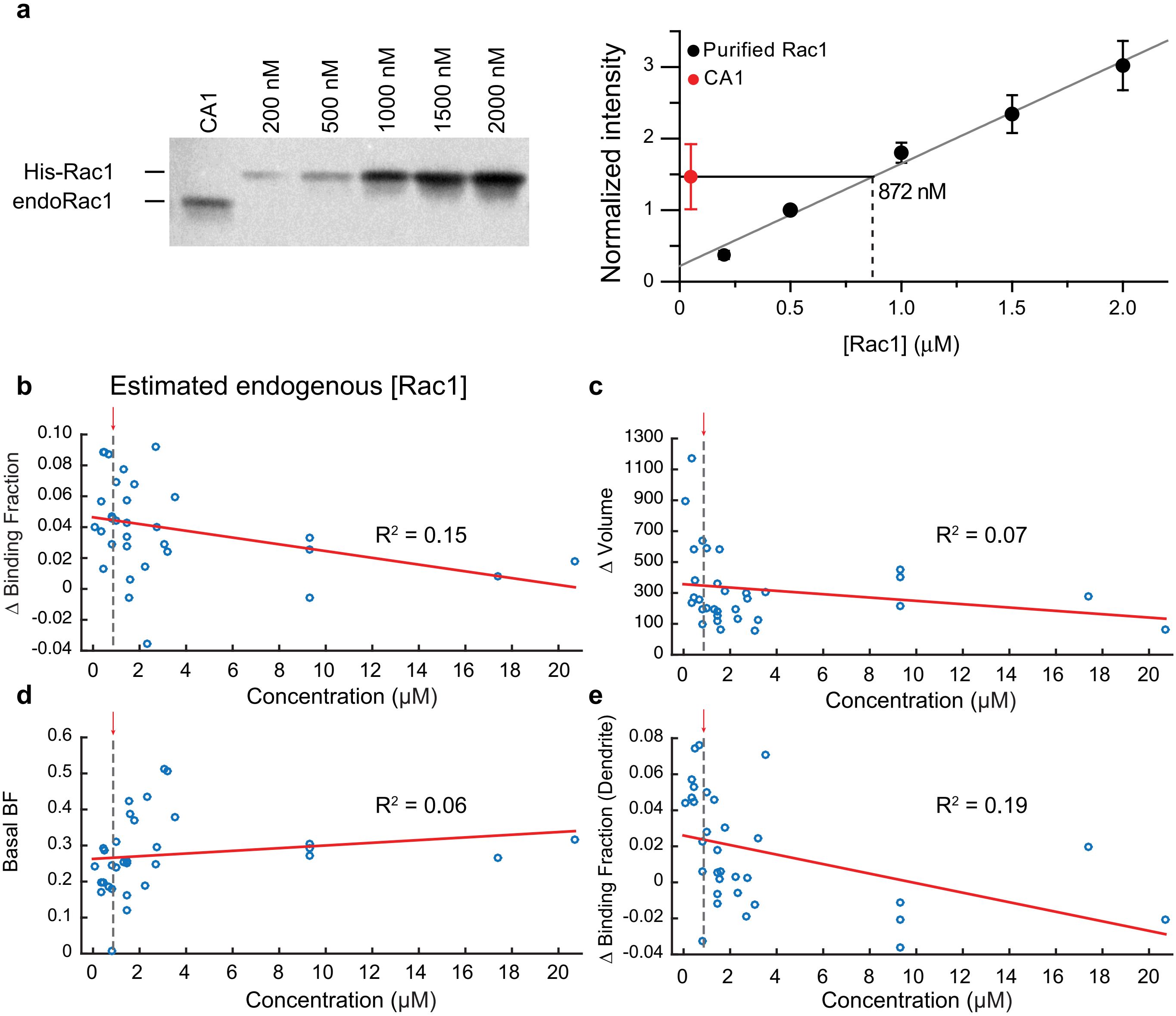
Extended Data Figure 1 | See next page for caption.

Extended Data Figure 1 | Design and characterization of the Rac1 sensor. a, Measurements of the affinity between sfGFP–Rac1 or sfGFP–Cdc42 and the p21-activated-kinase-derived acceptor construct, PAK2(65–117)^{R71C,S78A}. The binding fraction was measured using 2pFLIM (see Methods) across several concentrations of the acceptor construct. The dissociation constant was obtained by fitting the data (red) with a Michaelis–Menten function (grey). **b**. Representative fluorescence lifetime images of Rac1 sensor variants in HEK293T cells. Cells were transfected with a 1:2 donor:acceptor ratio of wild-type, dominant-negative (Rac1^{T17N}), substrate-binding dead (Rac1^{Y40C}), or constitutively active (Rac1^{Q61L}) variants of Rac1 with mCh–PBD2^{R71C,S78A}–mCh. Some experiments included the addition of an additional construct (the Rac1 guanine nucleotide exchange factor (GEF) Tiam1, or GTPase-activating protein (GAP) ARHGAP15) in a 1:2:1 donor:acceptor:GEF/GAP ratio. Cells were imaged 12–36 h after transfection in a warmed solution containing 30 mM Na-HEPES, pH 7.3, 130 mM NaCl, 2.5 mM KCl, 1 mM $CaCl_2$, 1 mM $MgCl_2$, 2 mM $NaHCO_3$, 1.25 mM NaH_2PO_4 and 25 mM glucose. Warmer colours indicate a lower fluorescence lifetime value/higher binding fraction of eGFP–Rac1 to the acceptor construct. Scale bars, 50 μm. **c**, Basal binding fraction of Rac1–PBD2$^{R71C, S78A}$ for the conditions listed in **b**. Error bars represent s.e.m. $*P < 0.0001$ compared to wild-type Rac1 control condition, ANOVA followed by post-hoc tests using the least significant difference. **d**, Time course of Rac1 activation in HEK293T cells upon application of 100 μg ml^{-1} epidermal growth factor (EGF). Control experiment corresponds to eGFP–Rac1 (donor) plus mCh–PBD2^{R71C,S78A}–mCh (acceptor) expression alone ($n = 39$ cells/8 plates), Rac1^{T17N} ($n = 33$ cells/4 plates) and Rac1^{Y40C} ($n = 40$ cells/5 plates) corresponds to the expression of the donor variant with the acceptor in the same ratio as controls, and Rac1plus W56 corresponds to the expression of the wild-type donor with the acceptor and the Rac1 inhibitory peptide W56.

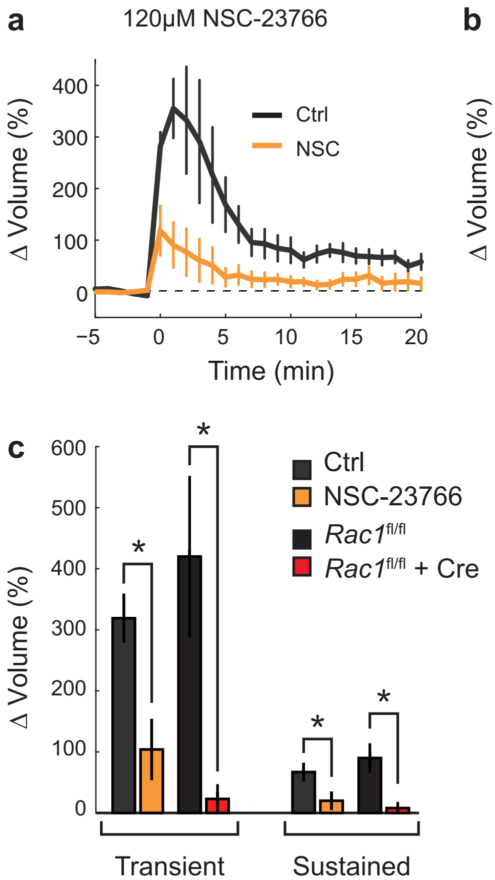


Extended Data Figure 2 | See next page for caption.

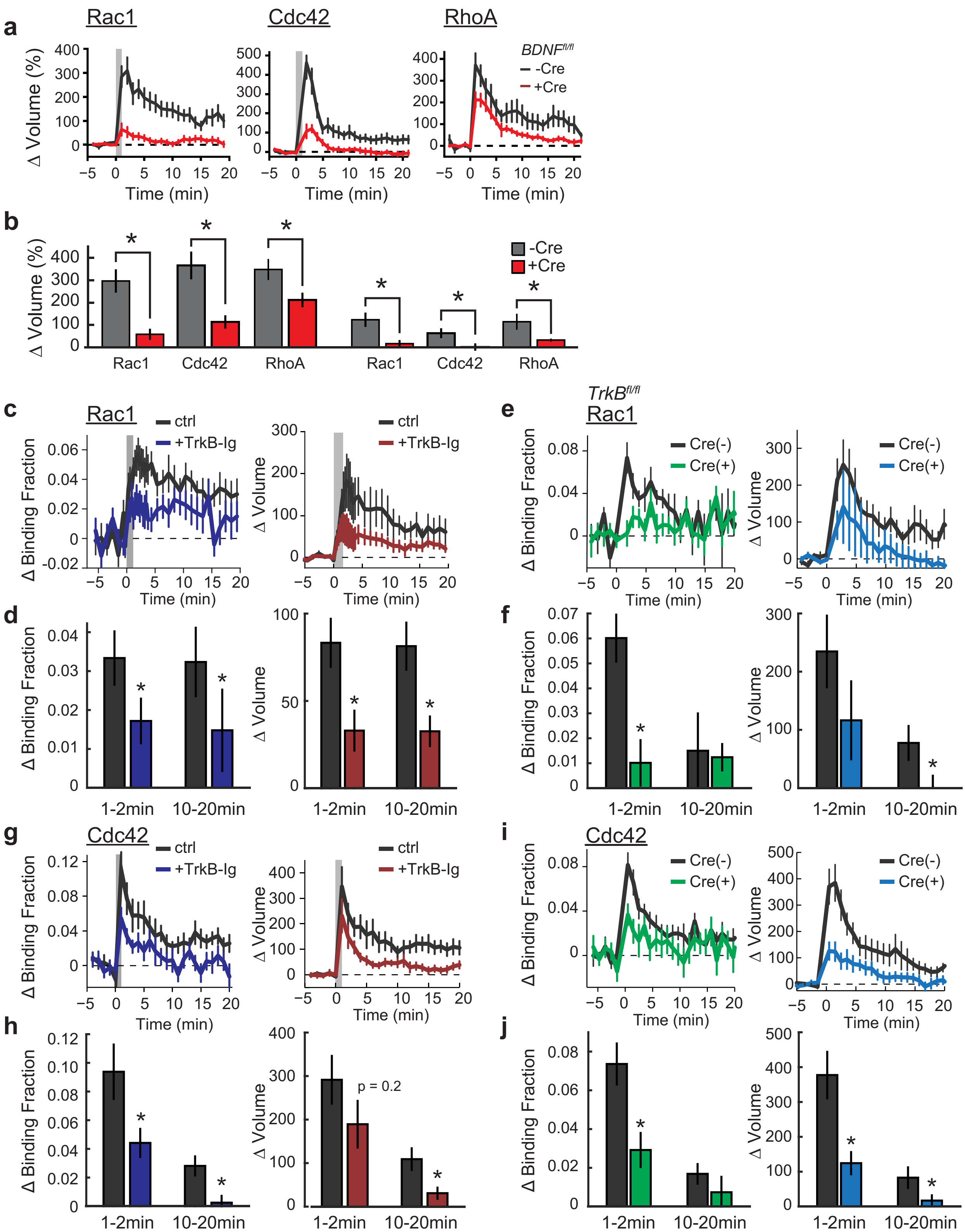
Extended Data Figure 2 | Characterization of the Rac1 sensor during single spine structural plasticity in hippocampal slices. a, Left, comparison of the signal spreading of Rac1, RhoA, Cdc42 and TrkB, normalized to spine activity from 1–2 min after stimulus onset. Error bars represent s.e.m. Data correspond to that shown in Fig. 1. Right, comparison of the spreading indices of the four sensors (spreading index = activity in dendrite (1–5 μm) from 1–2 min post-stimulation/maximum spine activation). Error was estimated by bootstrapping. $*P < 0.05$, independent-samples *t*-test. **b**, Same as **a**, but for 15–20 min after stimulation. **c**, Analysis of the diffusion time constants of constitutively active (Q61L) and wild-type Rac1 using photo-activatable GFP (paGFP). Left, representative 2p images of paGFP–Rac1^{Q61L} in a single dendritic spine after photo-activation. While active Rac1 shows a slightly slower diffusion than wild-type Rac1 (middle), both Rac1^{Q61L} and wild-type Rac1 diffuse away from the spine within approximately 10 s, similarly to Cdc42 and RhoA (ref. 3) (right two plots; wild type: $n = 41$ spines/4 cells; Rac1^{Q61L}: 67 spines/7 cells). **d**, Pharmacological characterization of Rac1 signal during sLTP. Rac1 activation in control ($n = 102$ cells/121 spines) conditions during 2p-glutamate uncaging, in the presence of the NMDAR blocker APV (100 μM; $n = 6$ cells/13 spines), and in the presence of the cell-permeable CaMKII inhibitory peptide tatCN21 (CN21; 10 μM; $n = 5$ cells/11 spines). Error bars represent s.e.m. **e**, Time course of spine volume change for experiments (**d**). Error bars represent s.e.m. **f**, Summary of effect of AP5 and CN21 on the transient (1–2 min after stimulation; control = 0.049 ± 0.003; AP5 = 0.00 ± 0.01; CN21 = 0.01 ± 0.01) and sustained (>10 min after stimulation; control = 0.033 ± 0.002; AP5 = 0.006 ± 0.007; CN21 = 0.026 ± 0.05) phases of Rac1 activation during sLTP. Error bars represent s.e.m. $*P < 0.05$, independent-samples *t*-test. **g**. Effects of near-physiological temperature of Rac1 sensor activation. Perfusion was warmed with a heating block holding the ACSF container, and the temperature measured at the back of the perfusion chamber (room temperature (RT), $n = 102$ cells/121 spines; for 30–32 °C, $n = 11$ cells/13 spines). Error bars represent s.e.m. **h**, Variability of unstimulated spine volume changes after the induction of sLTP in a nearby spine in Rac1 sensor-overexpressing neurons. Data shown are time courses of unstimulated spines close to the site of sLTP induction. While there is no average change in nearby spine volume with this stimulus (Fig. 1c), there is occasional enlargement or shrinkage. Data correspond to 100 randomly selected spines from the total of 777 nearby spines measured for the average depicted in Fig. 1c. **i**, Transient (1–2 min) change in volume of nearby spines as a function of distance from the stimulated spine. Data correspond to all nearby spines measured for the experiments in which the spatial profile of Rac1 was measured ($n = 56$ cells/79 experiments/218 nearby spines; Fig. 1e). Inset equation corresponds to the linear model of best fit. **j**, Same as **i**, but for the sustained (10–20 min) change in volume of nearby spines.



Extended Data Figure 3 | Estimation of endogenous Rac1 concentration and effects of Rac1 sensor overexpression. a, Left, representative western blot used to analyse endogenous Rac1 expression (endo-Rac1) from the CA1 region of hippocampal slices cultures compared to known concentrations of purified polyhistidine-tagged Rac1 (His-Rac1). Right, quantification of protein expression level for His-Rac1 from western blot shown in **a** averaged over three experiments. The concentration of endogenous Rac1 was estimated by measuring the intersection of the intensity of Rac1 protein from CA1 on the established calibration curve. **b**–**e**, Effect of concentration of the Rac1 sensor on the observed change in binding fraction or volume of the spine with glutamate uncaging (**b** or **b**, respectively), the basal binding fraction of the sensor (**d**), or the change in binding fraction in the dendrite in response to uncaging (**e**). Rac1 sensor concentration was estimated by making a standard curve of the intensity values known concentrations of eGFP at a range of imaging powers. The estimated endo-Rac1 concentration is plotted with a dashed line for comparison.

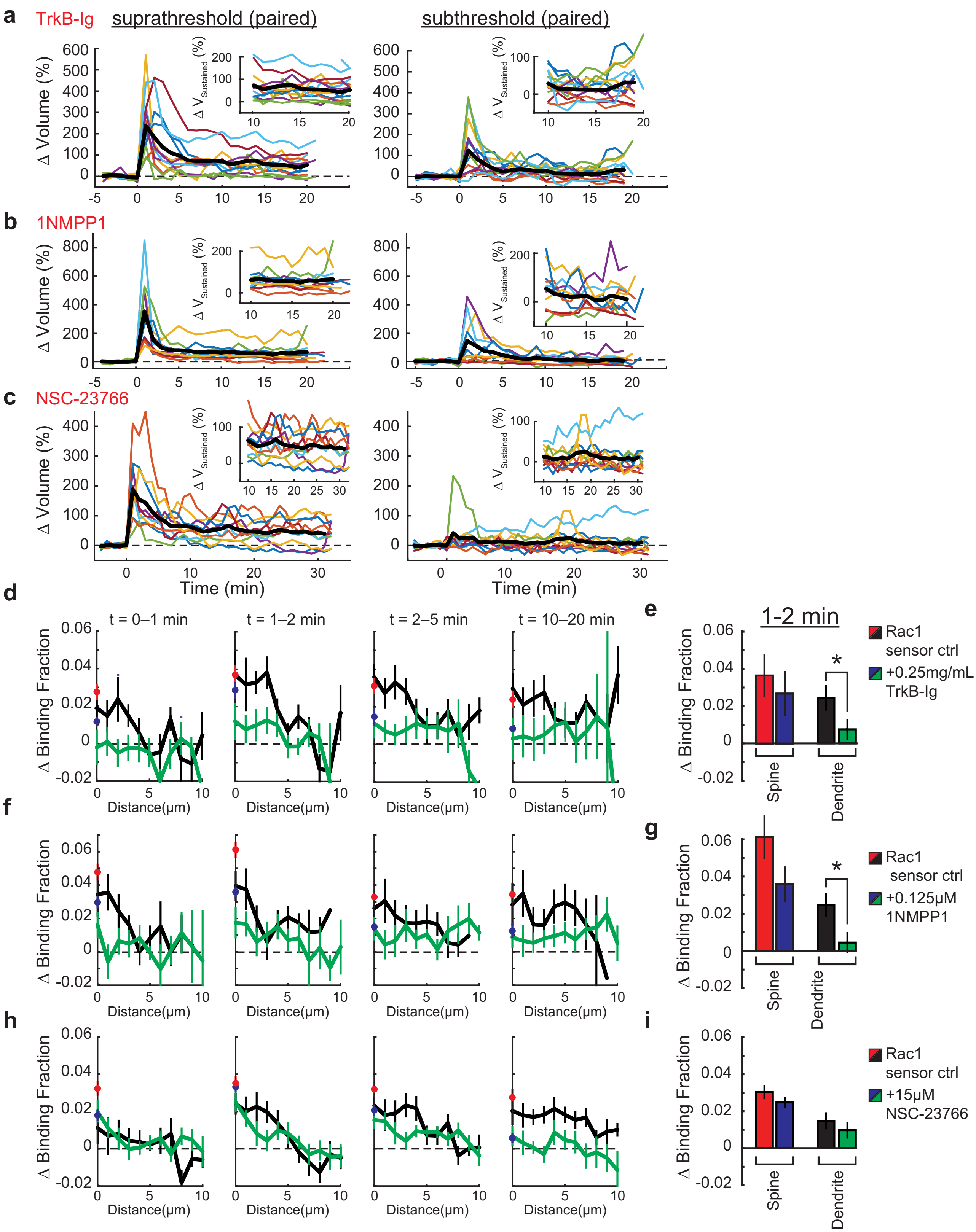


Extended Data Figure 4 | Effect of Rac1 inhibition on sLTP. a. The effect of a high concentration of NSC-23766 (120 µM) on uncaging-evoked spine volume change. Experiments were performed in eGFP-expressing rat hippocampal CA1 pyramidal neurons. Error bars represent s.e.m. $n = 5$ cells/9 control spines; $n = 5$ cells/15 NSC-23766 spines. **b**, Effect of single-cell Rac1 knockout on spine sLTP. *Rac1$^{fl/fl}$* slices were transfected with either eGFP alone (Cre (−); black curve, $n = 5$ cells/5 spines) or eGFP + tdTomato-Cre recombinase (Cre (+); red curve; $n = 7$ cells/7 spines). Error bars represent s.e.m. **c**, Summary of data in **a** and **b**. For NSC-23766 experiments, control $\Delta V_{transient} = 319 \pm 38\%$; control $\Delta V_{sustained} = 67 \pm 13\%$; NSC-23766 $\Delta V_{transient} = 104 \pm 48\%$; NSC-23766 $\Delta V_{sustained} = 20 \pm 13\%$. For Rac1 knockout experiments, Cre(−) $\Delta V_{transient} = 420 \pm 130\%$; Cre(−) $\Delta V_{sustained} = 90 \pm 22\%$; Cre(+) $\Delta V_{transient} = 23 \pm 9\%$; Cre(+) $\Delta V_{sustained} = 8 \pm 8\%$. Error bars represent s.e.m. $^*P < 0.05$, independent-samples *t*-test. All data are mean ± s.e.m.



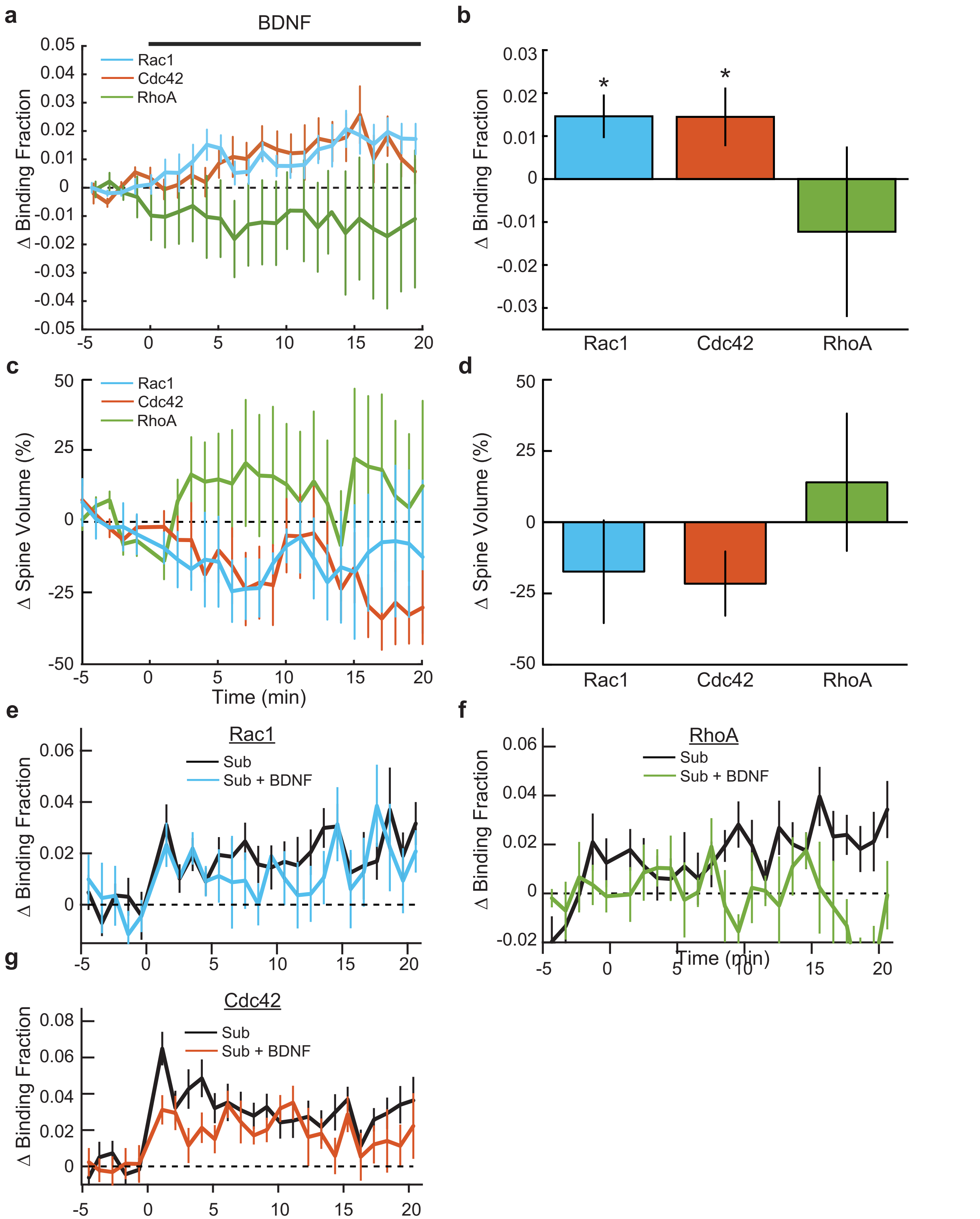
Extended Data Figure 5 | See next page for caption.

Extended Data Figure 5 | Characterization of Rho-GTPase dependence on BDNF–TrkB signalling. a, **b**, Dependence of sLTP (Δvolume) in Rho GTPase sensor-expressing neurons on postsynaptic BDNF. Data correspond to the volume data from experiments shown in Fig. 1f. Black denotes Cre^{-}, red denotes Cre^{+}. **b**, Summary of data from **a**. Bars represent the average spine volume change from 10 min after stimulation to the end of the experiment. **c**, **d**, Dependence of Rac1 on extracellular BDNF. **c**, Left, Rac1 (ctrl: $n=7$ cells/11 spines; TrkB-Ig: $n=8$ cells/14 spines), activity in the presence of 2 mg ml^{-1} TrkB-Ig. Right, spine volume change in Rac1 sensor-overexpressing cells in the two conditions. Grey bars indicate duration of uncaging bout. **d**, Summary of data in **c**. **e**, **f**, Dependence of Rac1 on postsynaptic TrkB. **e**, Left, Rac1 (Cre(−): $n=3$ cells/8 spines; Cre(+): $n=3$ cells/8 spines), activity in the presence or absence of Cre recombinase in *Trkb*$^{fl/fl}$ mouse slices. Right, spine volume change in Rac1 sensor-overexpressing cells in the two conditions. Grey bars indicate duration of uncaging bout. **f**, Summary of data in **e**. **g**, **h**, Dependence of Cdc42 on extracellular BDNF. Left, Cdc42 ($n=$ 7 cells/12 spines, 5 cells, 12 TrkB-Ig spines), activity in the presence of 2 mg ml^{-1} TrkB-Ig. Right, spine volume change in Rac1 sensor-overexpressing cells in the two conditions. Grey bars indicate duration of uncaging bout. **h**, Summary of data in **g**. **i**, **j**, Dependence of Cdc42 on postsynaptic TrkB. **i**, Left, Cdc42 (Cre(−): $n=5$ cells/12 spines; Cre(+): $n=6$ cells/16 spines), activity in the presence or absence of Cre recombinase in *Trkb*$^{fl/fl}$ mouse slices. Right, spine volume change in Cdc42 sensor-overexpressing cells in the two conditions. Grey bars indicate duration of uncaging bout. **j**, Summary of data in **i**. All data are mean ± s.e.m. $*P<0.05$, two-tailed independent-samples *t*-test.

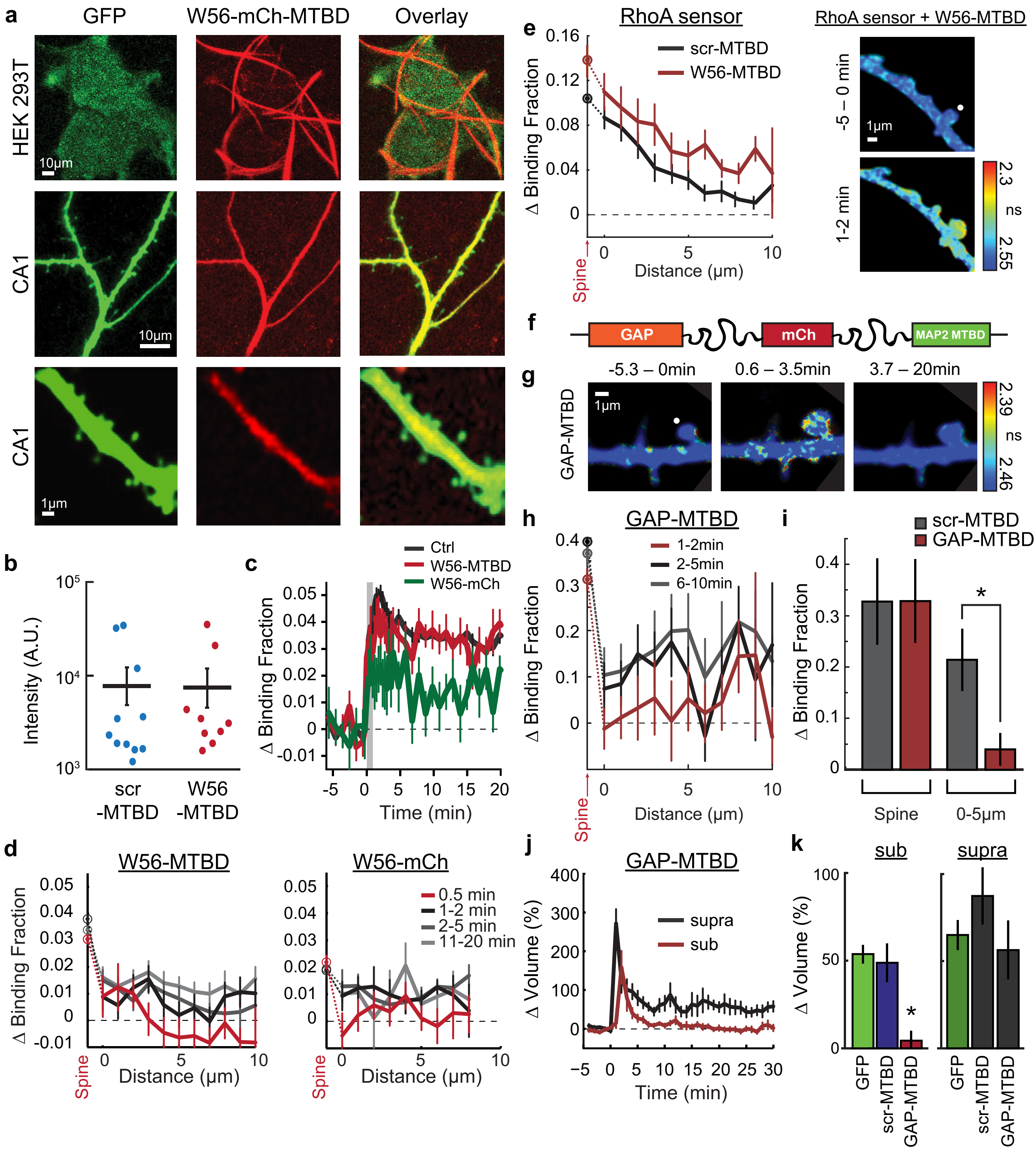


Extended Data Figure 6 | See next page for caption.

Extended Data Figure 6 | Characterization of the effects of weak pharmacological inhibition of BDNF–TrkB signalling on sLTP and Rac1 activity spreading. a. Individual data points for crosstalk experiments under low [TrkB-Ig] exposure. Each plot contains the average values of the experiments (thick, black curve) along with the corresponding individual experiments for the suprathreshold spine (left) and the subthreshold spine (right). Inset figures correspond to a close-up view of the data distributions for >10 min (the values used to calculate the sustained volume changes shown in Fig. 2). **b**, Same as **a**, but for crosstalk experiments using low [1NMPP1]. **c**, Same as **a** and **b**, but for crosstalk experiments using low [NSC-23766]. **d**, Effect of 0.125 μg ml^{-1} TrkB-Ig on Rac1 signal spreading. Each plot represents a specific time epoch after glutamate uncaging onset. Curves represent control (red = spine; black = dendrite) and +TrkB-Ig (blue = spine; green = dendrite) conditions plotted as a function of distance from the stimulated spine (n = 5 cells/6 control spines; n = 5 cells/9 +TrkB-Ig spines). **e**, Summary of data in **d**. Bars represent averages of the indicated temporal window across 1–5 μm of the dendrite. **f**, Same as **d**, but with the absence (red/black) and presence (blue/green) of 0.125 μM 1NMPP1 in *Trkb*^{F616A} slices (control: n = 5 cells/8 spines; +1NMPP1: n = 5 cells/11 spines). **g**, Summary of data in **f**. Bars represent averages of the indicated temporal window across 1–5 μm of the dendrite. **h**, Same as **d** and **f**, but in the absence (red/black) and presence (blue/green) of the Rac1 inhibitor, 15 μM NSC-23766 (control: n = 6 cells/8 spines; NSC-23766: n = 8 cells/13 spines). **i**, Summary of data in **h**. Bars represent averages of the indicated temporal window across 1–5 μm of the dendrite. *P < 0.05, t-test.

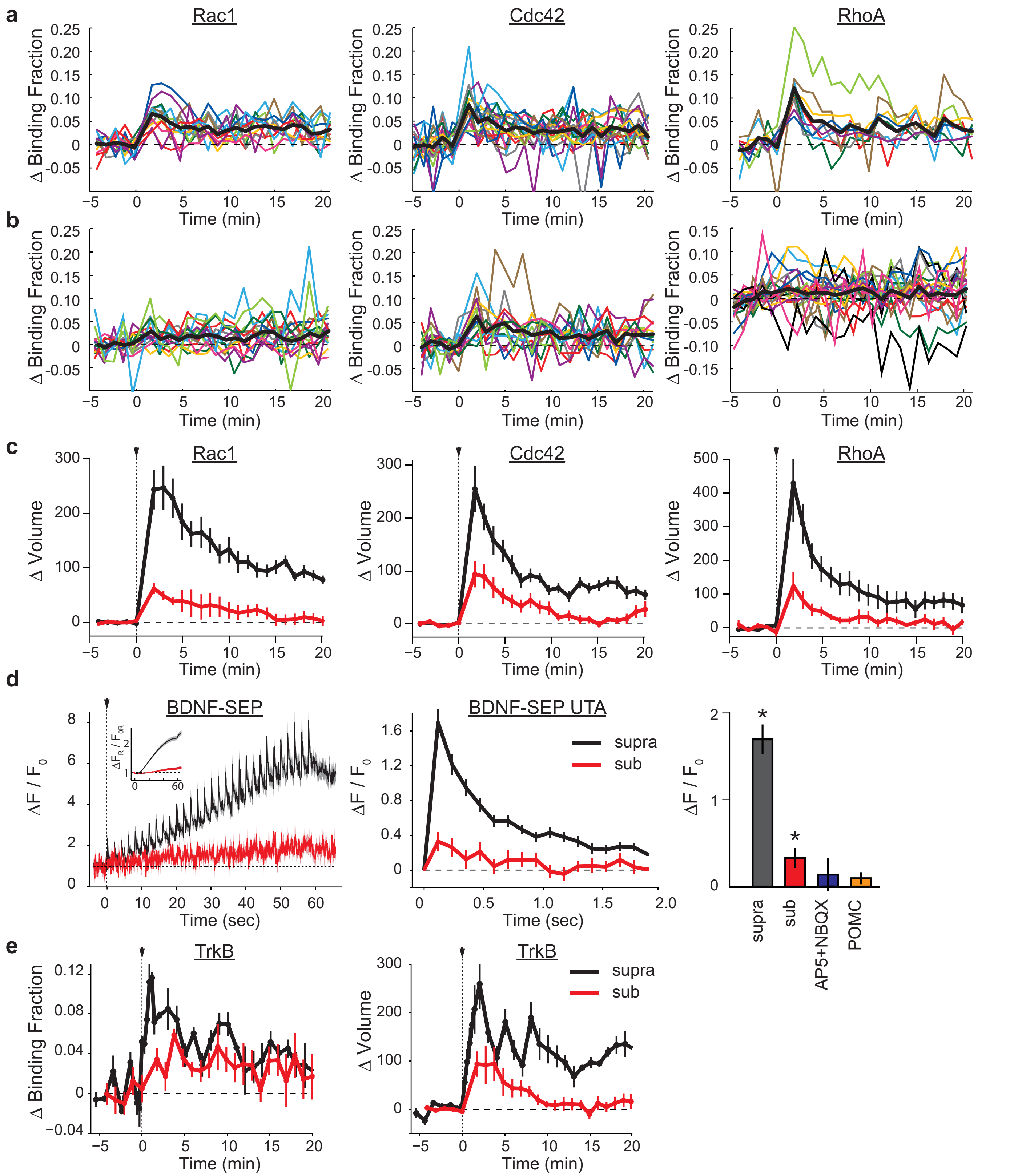


Extended Data Figure 7 | Effect of BDNF application on Rho GTPase signalling. a, Effect of bath application of 20 ng ml^{-1} exogenous BDNF on Rac1 ($n = 8$ cells), Cdc42 ($n = 10$ cells), and RhoA ($n = 5$ cells) sensors. $*P < 0.05$, statistically different from zero, one-sample t-test. **b**, Summary of data in **a**. **c**, Effects of BDNF application on spine volume in Rac1 (blue), Cdc42 (orange), and RhoA (green) sensor-expressing cells. Data are from the experiments in **a** and **b**. **d**, Summary of data in **c**. ANOVA determined that there was no significant difference between the sensor conditions. None of the conditions showed a significant difference from zero as determined by a one-sample t-test. **e**, Response of Rac1 to a subthreshold stimulus in the presence of 20 ng ml^{-1} BDNF (control $n = 15$ cells/15 spines; +BDNF $n = 6$ cells/6 spines) **f**, Same as **e**, but for RhoA (control $n = 14$ cells/18 spines; +BDNF $n = 7$ cells/8 spines) **g**, Same as **e** and **f**, but for Cdc42 (control $n = 17$ cells/20 spines; +BDNF $n = 7$ cells/7 spines).



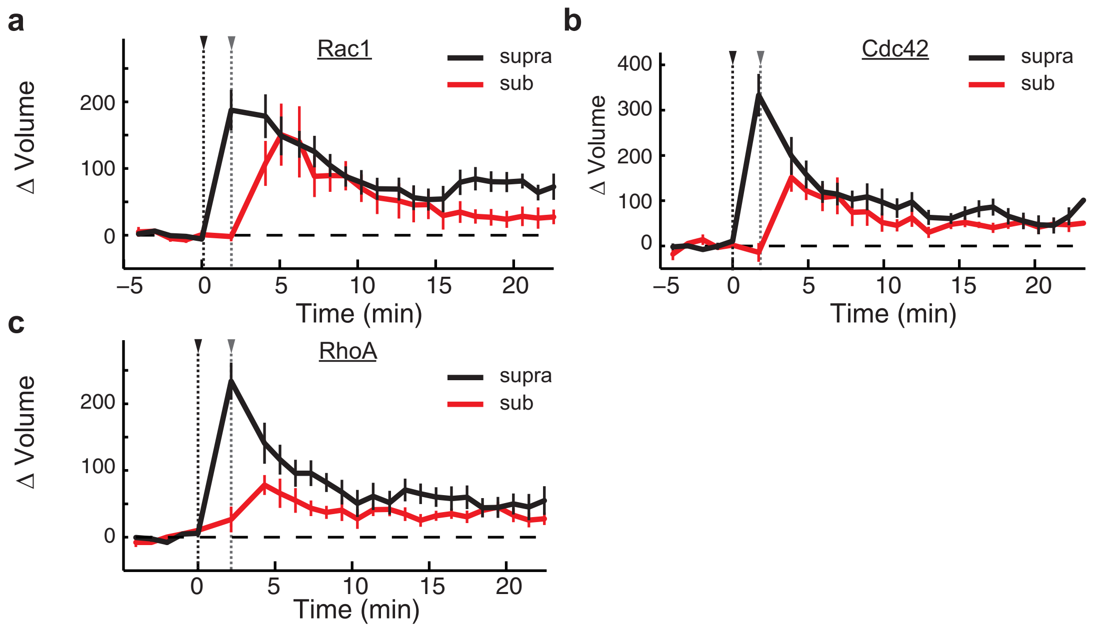
Extended Data Figure 8 | See next page for caption.

Extended Data Figure 8 | Characterization of the dendritic Rac1 inhibitor approach. a, Representative two-photon images illustrating the filamentous distribution HEK293T cells (top row), consistent with microtubule localization, and the largely dendrite-specific localization in CA1 neurons (bottom two rows) of W56–mCh–MTBD in comparison to an eGFP cell fill (first column). HEK293T cells were imaged at 1,050 nm to increase specificity of excitation for mCh versus eGFP. All images were acquired 2–5 days after transfection. **b**, Comparison of the expression levels of scr–MTBD and W56–MTBD from a subset of cells used for synaptic crosstalk experiments. Expression was estimated by acquiring intensity values in the red channel of an $\sim$10 μm section of a secondary dendrite using ImageJ. Each point represents a single cell. **c**, Average time course of Rac1 activation for cells expressing only the Rac1 sensor (control, black curve; $n = 105$ spines), the Rac1 sensor plus W56–mCh–MTBD (red curve; $n = 21$ cells/33 spines), and the Rac1 sensor plus W56–mC (green curve; $n = 6$ cells/13 spines). Error bars represent s.e.m. **d**, Comparison of the effect of W56–mCh–MTBD (left; $n = 21$ cells/33 spines) to untargeted W56–mCh (right; $n = 6$ cells/13 spines) on the spatial profile of Rac1 activation. Data are mean ± s.e.m. Control corresponds to the data in Fig. 1d. **e**, Effect of expression of W56–mCh–MTBD on RhoA sensor activation. Left, spatial profile of RhoA activation during sTLP induction in the presence of scr–mCh–MTBD control (black curve) and W56–mCh–MTBD (red curve). Right, representative 2pFLIM images of RhoA activation in the presence of W56–mCh–MTBD. White circle indicates the targeted spine. **f**, Schematic of the design of the GAP-based Rac1 dendritic inhibitor. A Rac1-specific GTPase-activating protein (ARHGAP15) replaced W56 in the general dendritic inhibitor construct (see Fig. 3a). **g**, Representative 2pFLIM images of the effect of GAP–mCh–MTBD on Rac1 signal spreading for the indicated time windows. White circle indicates the targeted spine. **h**, Quantification of the effects of expression of GAP–mCh–MTBD on Rac1 signal spreading after glutamate uncaging. Data are depicted as the change in binding fraction in the dendrite as a function of distance from the stimulated spine (change in binding fraction in spine plotted on *y* axis) ($n = 8$ cells/11 spines). **i**, Summary of the data depicted in **g** and **h**. Also shown are data for Rac1 spreading from scr–MTBD (see Fig. 3d) for comparison. $*P < 0.05$, independent-samples *t*-test. **j**. Effects of GAP–mCh–MTBD expression on synaptic crosstalk ($n = 4$ cells/8 crosstalk experiments). **k**, Summary of the data in **j**. $*P < 0.05$, ANOVA and post-hoc test using the Tukey–Kramer method.



Extended Data Figure 9 | See next page for caption.

Extended Data Figure 9 | Effect of a subthreshold stimulus on single spines of sensor-expressing CA1 neurons. a, Plots showing the variability of the response of the Rac1 (left), Cdc42 (middle), and RhoA (right) sensors to a suprathreshold stimulus. Thick black curves correspond to the averages depicted in Fig. 4a–c ('unpaired' condition) of the main text. The peak responses (1–2 min after stimulation, the same points used for statistical claims in Fig. 4) were subjected to a Shapiro–Wilk test to confirm the normality of the data. All data supported the null hypothesis, illustrating that the data are Guassian distributed and justifying the use of parametric statistics. **b**, Plots showing the variability of the response of the Rac1 (left), Cdc42 (middle) and RhoA (right) sensors to a subtreshold stimulus. Thick black curves correspond to the averages depicted in Fig. 4a–c ('unpaired' condition). **b**, Change in spine volume with glutamate uncaging (arrow and dotted line) during an unpaired threshold (black) and subthreshold (red) stimuli for CA1 pyramidal cells expressing the Rac1 (left), Cdc42 (centre), or RhoA (right) sensors. Shaded region represents s.e.m. Data correspond to the volume curves for the data presented in Fig. 4. **c**, 8-Hz two-photon imaging of BDNF–SEP intensity during an unpaired suprathreshold (black) or subthreshold (red) stimulus. Left, the full time course. Glutamate uncaging stimuli are delivered at 0.5 Hz beginning at $t = 0$, indicated by the black arrow and dashed line. Inset shows the associated volume change (as measured from mCh cell fill) of the two conditions. Middle panel, the uncaging-triggered average of 30 16-frame bins (corresponding to each uncaging pulse) and thus shows the average response to individual glutamate uncaging events. Right panel, the average of the first point after uncaging for the indicated conditions. Both suprathreshold and subthreshold conditions show a statistically significant difference from zero, while the presence of AP5 plus NBQX or the POMC peptide eliminate this signal. Error bars represent s.e.m. ($n = 28$ cells/217 spines (LTP), 5/84 (Sub), 2/46 AP5+NBQX), and 2/29 (POMC)). **d**, Left, activation of the TrkB sensor in response to an unpaired suprathreshold (black; $n = 4$ cells/5 spines) or subthreshold (red; $n = 6$ cells/8 spines) stimulus. Right, change in spine volume in a cell expressing the TrkB sensor in response to an unpaired threshold or subthreshold stimulus. Error bars represent s.e.m.



Extended Data Figure 10 | Change in volume of paired spines during synaptic crosstalk in Rho GTPase-expressing CA1 neurons. a, Spine volume change in response to a suprathreshold (black curve; black arrow indicates stimulus initiation) and a paired subthreshold (red curve, grey arrow indicates stimulus initiation) stimulus in spines from cells expressing the Rac1 sensor ($n = 6$ cells/12 spine pairs). Error bars represent s.e.m. **b**, Same as **a**, but for Cdc42-expressing cells ($n = 9$ cells/10 spine pairs). **c**, Same as **a** and **b**, but for RhoA-expressing cells ($n = 8$ cells/12 spine pairs).

LETTER

doi:10.1038/nature19788

The lipolysis pathway sustains normal and transformed stem cells in adult *Drosophila*

Shree Ram Singh[1]*, Xiankun Zeng[1]*, Jiangsha Zhao[1], Ying Liu[1], Gerald Hou[1], Hanhan Liu[1] & Steven X. Hou[1]

Cancer stem cells (CSCs) may be responsible for tumour dormancy, relapse and the eventual death of most cancer patients[1]. In addition, these cells are usually resistant to cytotoxic conditions. However, very little is known about the biology behind this resistance to therapeutics. Here we investigated stem-cell death in the digestive system of adult *Drosophila melanogaster*. We found that knockdown of the coat protein complex I (COPI)–Arf79F (also known as Arf1) complex selectively killed normal and transformed stem cells through necrosis, by attenuating the lipolysis pathway, but spared differentiated cells. The dying stem cells were engulfed by neighbouring differentiated cells through a draper–myoblast city–Rac1–basket (also known as JNK)-dependent autophagy pathway. Furthermore, Arf1 inhibitors reduced CSCs in human cancer cell lines. Thus, normal or cancer stem cells may rely primarily on lipid reserves for energy, in such a way that blocking lipolysis starves them to death. This finding may lead to new therapies that could help to eliminate CSCs in human cancers.

To investigate the molecular mechanism behind the resistance of CSCs to therapeutics, we studied the death of stem cells with different degrees of quiescence in the adult *Drosophila* digestive system, including intestinal stem cells (ISCs)[2,3], renal and nephric stem cells (RNSCs)[4] and hindgut intestinal stem cells (HISCs)[5,6] (Fig. 1a and Extended Data Fig. 1a). We found that expression of the proapoptotic genes *rpr* and *p53* effectively ablated differentiated cells but had little effect on stem cells (Extended Data Fig. 1b–n).

In mammals, treatment-resistant leukaemic stem cells (LSCs) can be eliminated by a two-step protocol involving initial activation by interferon-α (IFNα) or colony-stimulating factor (G-CSF), followed by targeted chemotherapy[7]. In *Drosophila*, activation of the hopscotch (also known as JAK)–Stat92E signalling pathway induces hyperplastic stem cells, which are overproliferating, but retain their apico-basal polarity and differentiation ability[4,6,8]. We conducted a slightly different two-step protocol in *Drosophila* stem cells by overexpressing the JAK–Stat92E pathway ligand unpaired (upd) and rpr together. The induction of *upd* + *rpr* using the temperature-sensitive (ts) mutant *esg*–*Gal4* ($esg^{ts} > upd + rpr$; Fig. 1b, c, j and Extended Data Fig. 1o–q) effectively ablated all of the ISCs and RNSCs through apoptosis within four days. Consistent with this result, expressing a gain-of-function *Raf* mutant (Raf^{gof}) also accelerated apoptotic cell death of hyperplastic ISCs[9].

Expressing a constitutively active form of Ras oncogene at 85D (also known as Ras^{V12}) in RNSCs[10] and the knockdown of Notch activity in ISCs[11,12] can transform these cell types into CSC-like neoplastic

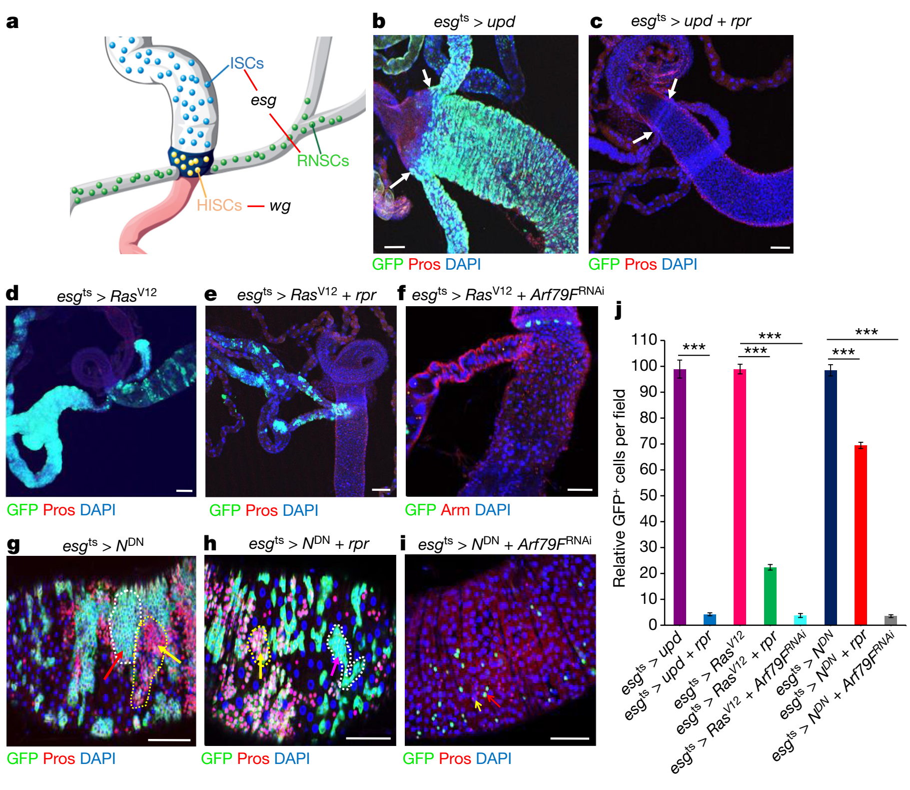


Figure 1 | Activation of proliferation accelerates apoptotic cell death of hyperplastic stem cells but fails to completely eliminate neoplastic stem cells. a, Diagram of three types of stem cells near the hindgut–midgut junction, and the cells in which *esg*–*Gal4* (*esg*) and *wg*–*Gal4* (*wg*) were expressed. **b–i,** Representative images of the posterior midguts of flies with the indicated phenotypes. **b,** $esg^{ts} > upd$, 29 °C, 4 d ($n = 33$). **c,** $esg^{ts} > upd + rpr$, 29 °C, 4 d ($n = 35$). **d,** $esg^{ts} > Ras^{V12}$, 29 °C, 7 d ($n = 34$). **e,** $esg^{ts} > Ras^{V12} + rpr$, 29 °C, 7 d ($n = 29$). **f,** $esg^{ts} > Ras^{V12} + Arf79F^{RNAi}$, 29 °C, 7 d ($n = 35$). **g,** $esg^{ts} > N^{DN}$, 29 °C, 7 d ($n = 24$). **h,** $esg^{ts} > N^{DN} + rpr$, 29 °C, 7 d ($n = 29$). **i,** $esg^{ts} > N^{DN} + Arf79F^{RNAi}$, 29 °C, 7 d ($n = 38$). **j,** Quantification of GFP^{+} cells from midguts isolated from flies with the indicated genotypes. Data are represented as mean ± s.e.m. Statistical significance determined by Student's *t*-test, $***P < 0.0001$. The posterior midguts of flies with the indicated genotypes were dissected, stained with the GFP and Prospero (Pros) antibodies and analysed by confocal microscopy. White arrows in **b** and **c** point to the hindgut–midgut junction. **g, h,** Red arrows with white dotted lines point to clusters of ISCs enteroblasts and yellow arrows with yellow dotted lines point to clusters of enteroendocrine cells. **i,** Red and yellow arrows point to remaining ISCs/enteroblasts, and enteroendocrine cells, respectively. Scale bars in **b–i**, 10 μm.

[1]The Basic Research Laboratory, National Cancer Institute at Frederick, National Institutes of Health, Frederick, Maryland 21702, USA.
*These authors contributed equally to this work.

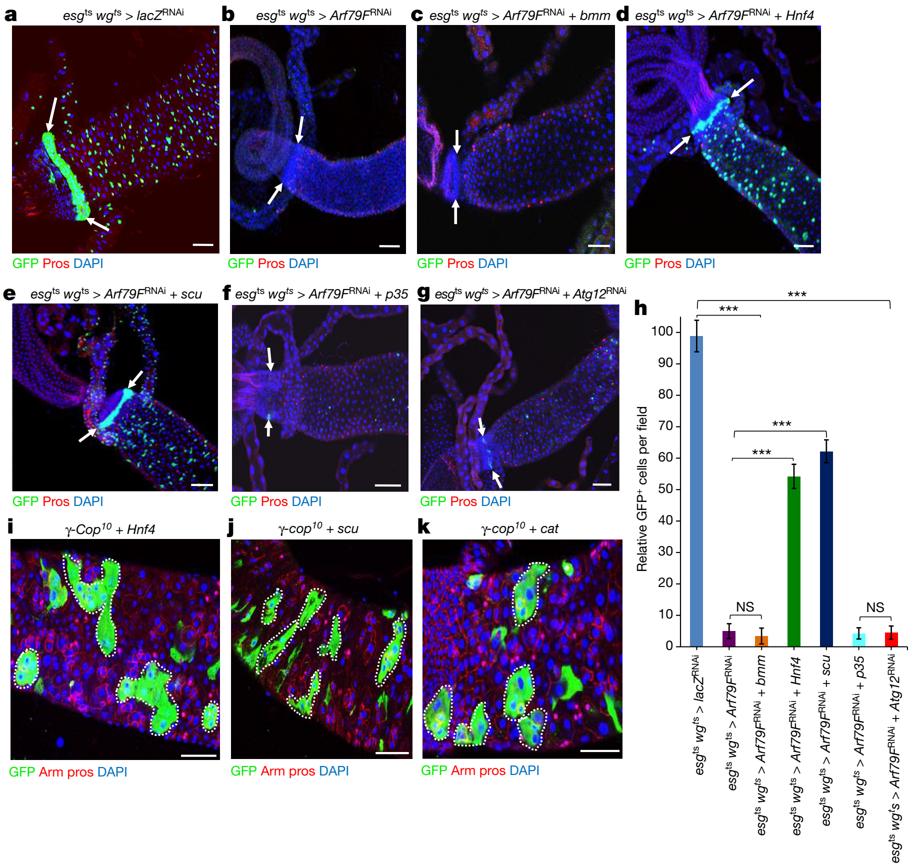


Figure 2 | The COPI–Arf79F complex regulates stem cell survival through a lipolysis pathway. **a**–**g**, Representative images are shown. The genotypes of the flies in each panel were: **a**, esg^{ts} wg^{ts} > $lacZ^{RNAi}$, 29 °C, 7 d (n = 34). **b**, esg^{ts} wg^{ts} > $Arf79F^{RNAi}$, 29 °C, 7 d (n = 38). **c**, esg^{ts} wg^{ts} > $Arf79F^{RNAi}$ + *bmm*, 29 °C, 7 d (n = 31). **d**, esg^{ts} wg^{ts} > $Arf79F^{RNAi}$ + *Hnf4*, 29 °C, 7 d (n = 29). **e**, esg^{ts} wg^{ts} > $Arf79F^{RNAi}$ + *scu*, 29 °C, 7 d (n = 32). **f**, esg^{ts} wg^{ts} > $Arf79F^{RNAi}$ + *p35*, 29 °C, 7 d (n = 37). **g**, esg^{ts} wg^{ts} > $Arf79F^{RNAi}$ + $Atg12^{RNAi}$, 29 °C, 7 d (n = 32). **h**, Quantification of GFP^{+} cells from midguts isolated from flies with the indicated genotypes. Data are represented as mean ± s.d. Statistical significance determined by Student's *t*-test, ***$P < 0.0001$. NS, not significant ($P > 0.05$). **i**–**k**, MARCM clones of flies with the following genotypes: **i**, *UAS–Hnf4; FRT^{82B}–γ-Cop^{10}*, 7 d after clonal induction (ACI). **j**, *UAS-scu; FRT^{82B}-γ-Cop^{10}*, 7 d ACI (n = 32). **k**, *UAS–cat; FRT^{82B}–γ-Cop^{10}*, 7 d ACI (n = 35). The posterior midguts of flies with the indicated genotypes were dissected, stained with the GFP, Prospero (Pros) and Armadillo (Arm) antibodies and analysed by confocal microscopy. White arrows in **a**–**g** point to the hindgut–midgut junction. White dotted lines in **i**–**k** outline GFP^{+} clones. Scale bars in **a**–**g** and **i**–**k**, 10 μm.

stem cells, which were not only overproliferating, but also lost their apico-basal polarity and differentiation ability (Fig. 1d, g). We found that expressing *rpr* in Ras^{V12}-transformed RNSCs (esg^{ts} > Ras^{V12} + *rpr*; Fig. 1e, j) or in ISCs expressing a dominant–negative form of Notch (N^{DN}) (esg^{ts} > N^{DN} + *rpr*; Fig. 1h, j) caused the ablation of only a proportion of the transformed RNSCs and few transformed ISCs and it did not affect differentiated cells (Extended Data Fig. 1r–u); substantial populations of the neoplastic stem cells remained even seven days after *rpr* induction.

These results suggest that the activation of proliferation can accelerate the apoptotic cell death of hyperplastic stem cells, but that a proportion of actively proliferating neoplastic RNSCs and ISCs are resistant to apoptotic cell death. Neoplastic tumours in *Drosophila* are more similar to high-grade malignant human tumours than are the hyperplastic *Drosophila* tumours[13].

Vesicle-mediated COPI and COPII are essential components of the trafficking machinery for vesicle transportation between the endoplasmic reticulum and the Golgi[14]. In addition, the COPI complex regulates the transport of lipolysis enzymes to the surface of lipid droplets for lipid droplet usage[15] (Extended Data Fig. 2a). In our previous screen, we found that knockdown of COPI components (including Arf79F, the *Drosophila* homologue of ADP-ribosylation factor 1 (Arf1)) rather than COPII components[16] resulted in stem-cell death, suggesting that lipid-droplet usage (lipolysis) rather than the general trafficking machinery between the endoplasmic reticulum and Golgi is important for stem-cell survival.

To further investigate the roles of these genes in stem cells, we used a recombined double Gal4 line of *esg–Gal4* and *wg–Gal4* to express genes in ISCs, RNSCs, and HISCs (esg^{ts} wg^{ts} > *X*). Knockdown of these genes using RNA interference (RNAi) in stem cells (esg^{ts} wg^{ts} > X^{RNAi}; Extended Data Fig. 2b–k) ablated most of the stem cells in 1 week. However, expressing $Arf79F^{RNAi}$ in enterocytes ($NP1^{ts}$ > $Arf79F^{RNAi}$; Extended Data Fig. 2l–o) or in differentiated stellate cells in Malpighian tubules (tsh^{ts} > $Arf79F^{RNAi}$; Extended Data Fig. 2p, q) did not cause similar marked ablation. These results suggest that Arf79F knockdown selectively kills stem cells and not differentiated cells.

We also found that expressing $Arf79F^{RNAi}$ (esg^{ts} > Ras^{V12} + $Arf79F^{RNAi}$; Fig. 1f, j) or ζ-COP^{RNAi} (esg^{ts} wg^{ts} >Ras^{V12} + ζ-COP^{RNAi}; Extended Data Fig. 2r) in Ras^{V12}-transformed RNSCs ablated almost all of the transformed stem cells. Similarly, expressing $Arf79F^{RNAi}$ (esg^{ts} > N^{DN} + $Arf79F^{RNAi}$; Fig. 1i, j) or δ-COP^{RNAi} (esg^{ts} wg^{ts} > N^{DN} + δ-COP^{RNAi}; Extended Data Fig. 2s) in N^{DN}-transformed ISCs ablated all of the cells within one week, but restored differentiated cells to close to their normal levels within one week (Extended Data Fig. 2t, u).

We further generated *δ-COP*- and *γ-COP*-mutant clones using the mosaic analysis with a repressible cell marker (MARCM) technique[17] and found that the COPI complex cell-autonomously regulated stem cell survival (Extended Data Fig. 3a–h). In summary, knockdown of the COPI–Arf79F complex effectively ablated normal and transformed stem cells but not differentiated enterocytes or stellate cells.

In our RNAi screen we also identified acyl-CoA synthetase long-chain (ACSL), an enzyme in the *Drosophila* lipolysis–β-oxidation pathway[16,18] (Extended Data Fig. 2a), and bubblegum (bgm), a very long-chain fatty acid-CoA ligase[16,19]. RNAi-mediated knockdown of *Acsl* (esg^{ts} wg^{ts} > $Acsl^{RNAi}$; Extended Data Fig. 2i, k) and *bgm* (esg^{ts} wg^{ts} > bgm^{RNAi}; Extended Data Fig. 2j, k) effectively killed ISCs and RNSCs, but killed HISCs less effectively. Expressing $Acsl^{RNAi}$ in Ras^{V12}-transformed RNSCs (esg^{ts} wg^{ts} > Ras^{V12} + $Acsl^{RNAi}$; Extended Data Fig. 2v) also ablated almost all of the transformed RNSCs in one week.

Brummer (bmm) is a triglyceride lipase, the *Drosophila* homologue of mammalian ATGL, the first enzyme in the lipolysis pathway[20] (Extended Data Fig. 3a). Scully (scu) is the *Drosophila* orthologue of hydroxy-acyl-CoA dehydrogenase, an enzyme in the β-oxidation pathway[21]. Hepatocyte nuclear factor 4 (Hnf4) regulates the expression of several genes involved in lipid mobilization and β-oxidation[21]. To determine whether the lipolysis–β-oxidation pathway is required for COPI–Arf79F-mediated stem cell survival, we expressed upstream

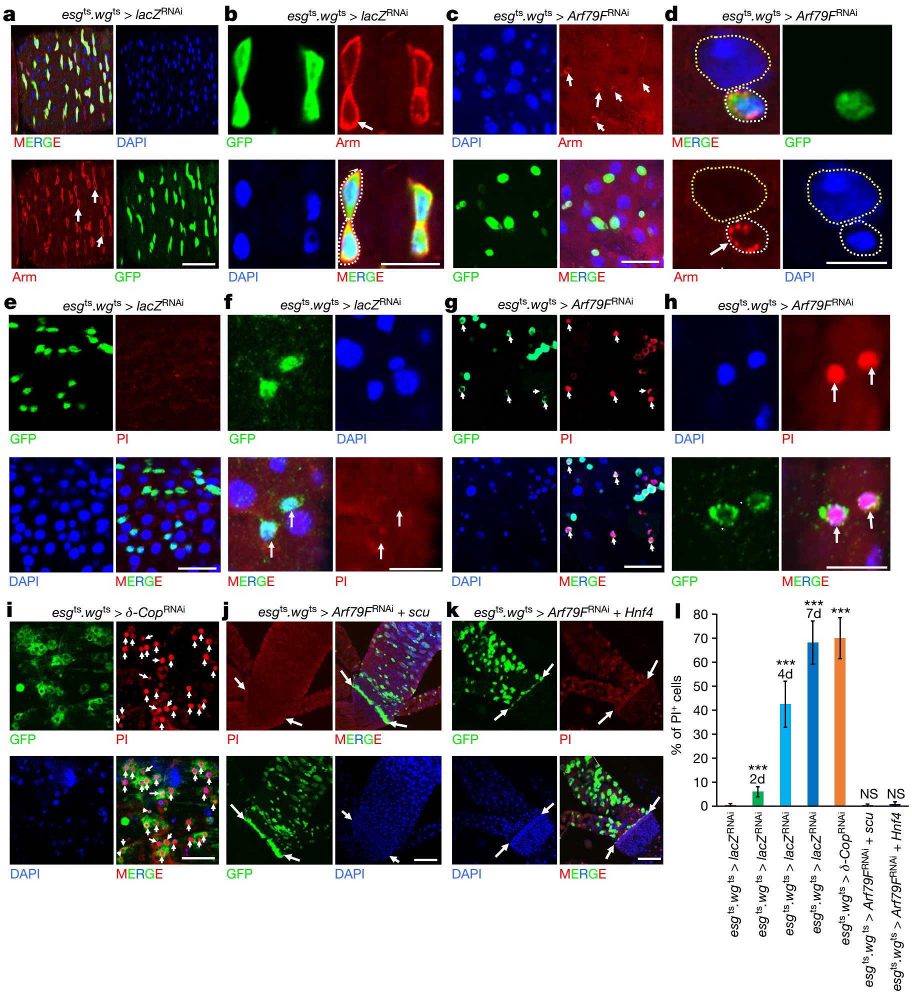


Figure 3 | Knockdown of components of the COPI–Arf79F–β-oxidation pathway kill stem cells through necrosis. The genotypes of the flies in each panel were: **a**, **b**, **e**, **f**, *esg*^{ts} *wg*^{ts} > *lacZ*^{RNAi}, 29 °C, 4 d (n = 32). **c**, **d**, **g**, **h**, *esg*^{ts} *wg*^{ts} > *Arf79F*^{RNAi}, 29 °C, 4 d (n = 28). **i**, *esg*^{ts} *wg*^{ts} > *δ-Cop*^{RNAi}, 29 °C, 4 d (n = 30). **j**, *esg*^{ts} *wg*^{ts} > *Arf79F*^{RNAi} + *scu*, 29 °C, 4 d (n = 28). **k**, *esg*^{ts} *wg*^{ts} > *Arf79F*^{RNAi} + *Hnf4*, 29 °C, 4 d (n = 30). **l,** Quantification of PI$^+$ cells from midguts isolated from flies with the indicated genotypes. Data are represented as mean ± s.d. Statistical significance determined by Student's *t*-test, ***P < 0.0001; NS, not significant (P > 0.05). The posterior midguts of flies with the indicated genotypes were dissected, stained with GFP and Armadillo (Arm) antibodies or indicated reagents, and analysed by confocal microscopy. White arrows in **a**–**i** point to esg–GFP$^+$ cells and in **j**–**k** point to the hindgut–midgut junction. White dotted lines in **d** outline ISCs/enteroblasts, yellow dotted lines outline enterocytes. Scale bars in **a**–**k**, 10 μm.

activating sequence (UAS)-regulated constructs (*UAS–bmm* (Fig. 2c, h), *UAS–Hnf4* (Fig. 2d, h), and *UAS–scu* (Fig. 2e, h)) in stem cells that were depleted of Arf79F (Fig. 2b–e), β-COP (Extended Data Fig. 2w), or ζ-COP (Extended Data Fig. 2x). Overexpressing either *scu* or *Hnf4* significantly (P < 0.0001) attenuated the stem cell death caused by knockdown of the COPI–Arf79F complex. Expressing *UAS–Hnf4* (Fig. 2i) and *UAS–scu* (Fig. 2j) in *FRT*^{82B}–*γ-COP*^{10} MARCM clones also rescued the stem cell death phenotype induced by γ-COP knockdown (Extended Data Fig. 3f, h). However, *bmm* overexpression did not rescue the stem-cell death induced by Arf79F knockdown (Fig. 2c, h). Since there are several other triglyceride lipases in *Drosophila* in addition to bmm, another lipase may redundantly regulate the lipolysis pathway.

To further investigate the function of lipolysis in stem cells, we investigated the expression of a lipolysis reporter (*GAL4–dHFN4; UAS–nlacZ*)[21], which consisted of *hsp70–GAL4–dHNF4* combined with a *UAS–nlacZ* reporter gene. The flies were either cultured continuously at 29 °C or heat-shocked for 30 min at 37 °C, 12 h before dissection. Without heat shock, the reporter was expressed only in ISCs and RNSCs of mature adult flies, but not in enteroendocrine cells, enterocytes, quiescent HISCs or quiescent ISCs of freshly emerged young adult flies (less than 3 days old) (Extended Data Fig. 3i–m). Expressing *δ-COP*^{RNAi} (*esg*^{ts} > *δ-COP*^{RNAi} + *GAL4–dHFN4; UAS–nlacZ*) almost completely eliminated the reporter expression (Extended Data Fig. 3n), suggesting that the reporter was specifically regulated by the COPI complex. After heat shock or when a constitutively active form of JAK (*hop*$^{Tum-l}$) was expressed, the reporter was strongly expressed in ISCs, RNSCs and HISCs, but not in enteroendocrine cells or enterocytes (Extended Data Fig. 3o, p). These data suggest that COPI-complex-regulated lipolysis was active in stem cells, but not in differentiated cells, and that the absence of the reporter expression in quiescent HISCs at 29 °C was probably owing to weak *hsp70* promoter activity rather than to low lipolysis in these cells.

We further investigated lipid storage, and found that the size and number of lipid droplets were markedly increased in stem cells after knockdown of *Arf79F* (*esg*^{ts} > *Arf79F*^{RNAi}) (Extended Data Fig. 3q–v).

We also used Arf1 inhibitors (brefeldin A, golgicide A, secin H3, LM11 and LG8) and fatty-acid-oxidation (FAO) inhibitors (triacsin C, mildronate, etomoxir and enoximone) and found that these inhibitors markedly reduced stem-cell tumours in *Drosophila* through the lipolysis pathway but had a negligible effect on normal stem cells (Extended Data Fig. 4).

These data together suggest that the COPI–Arf1 complex regulates stem-cell survival through the lipolysis–β-oxidation pathway, and that

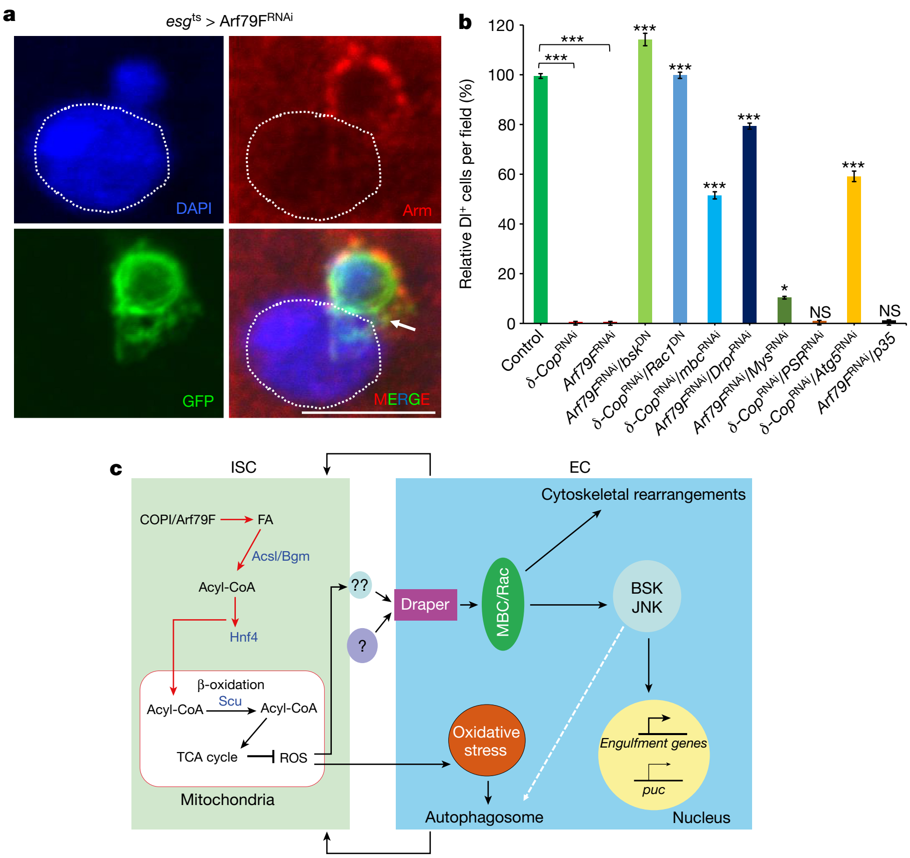


Figure 4 | Dying ISCs are engulfed by neighbouring enterocytes through the draper–Rac–JNK (Bsk) pathway. a, Representative images from $esg^{ts} > Arf79F^{RNAi}$ cells, 29°C, 7 d ($n = 32$). A dying ISC is engulfed by a neighbouring enterocyte. The posterior midguts of flies were dissected, stained with the GFP and Armadillo (Arm) antibodies, and analysed by confocal microscopy. Arrow points to GFP^{+} stem cell/progenitors and the dotted line outlines a nucleus and cell membrane of an enterocyte. Scale bar, 10 μm. **b**, Quantification of Dl^{+} ($Delta^{+}$) ISCs cells in flies with the $NP1^{ts}$ esg^{ts} driver and indicated genotypes (See images in Extended Data Fig. 9 for details). Data are represented as mean ± s.e.m. Statistical significance determined by Student's *t*-test, $*P < 0.05$, $***P < 0.0001$. NS, not significant ($P > 0.05$). **c**, Model of ISC death induced by knockdown of the COPI–Arf1 complex. Details are described in the text. The autophagosome is involved in the last step of phagocytosis (degradation of internalized cargo)[27], and autophagy may both function downstream of and be regulated by the drpr–Mbc–Rac1–JNK pathway.

knockdown of these genes blocks lipolysis but promotes lipid storage. Further, the transformed stem cells are more sensitive to Arf1 inhibitors and may be selectively eliminated by controlling the concentration of Arf1 inhibitors.

Our data suggest that neither caspase-mediated apoptosis nor autophagy-regulated cell death regulates the stem-cell death induced by the knockdown of components of the COPI–Arf79F complex (Fig. 2f–h). We therefore investigated whether necrosis regulates the stem-cell death induced by knockdown of the COPI–Arf79F complex. Necrosis is characterized by early plasma membrane rupture, reactive oxygen species (ROS) accumulation and intracellular acidification[22]. Propidium iodide detects necrotic cells with compromised membrane integrity, the oxidant-sensitive dye dihydroethidium (DHE) indicates cellular ROS levels and LysoTracker staining detects intracellular acidification[22]. We detected the membrane rupture phenotype only in $esg^{ts}.wg^{ts} > Arf79F^{RNAi}$ ISCs but not in wild-type ISCs (Figs 3a–d, 4a and Extended Data Fig. 5a–i) and the propidium iodide signal was observed only in ISCs from flies that had RNAi-induced knockdown of expression of COPI–Arf79F components (esg^{ts} $wg^{ts} > X^{RNAi}$, Fig. 3g–i, l and Extended Data Fig. 5k–p), and not in cells from wild-type (Fig. 3e, f, l and Extended Data Fig. 5j, p), *scu*-rescued $Arf79F^{RNAi}$ (esg^{ts} $wg^{ts} > scu + Arf79F^{RNAi}$, Fig. 3j, l) or *Hnf4*-rescued $Arf79F^{RNAi}$ (esg^{ts} $wg^{ts} > Hnf4 + Arf79F^{RNAi}$, Fig. 3k, l) flies. In the esg^{ts} $wg^{ts} > Acsl^{RNAi}$ flies, all of the ISCs and RNSCs were ablated after four days at 29°C, but a fraction of the HISCs remained, and these were also propidium iodide positive (Extended data Fig. 5n–p), indicating that the HISCs were dying slowly. This slowness may have been due to either a lower GAL4 (*wg–Gal4*) activity in these cells compared to ISCs and RNSCs (*esg–Gal4*) or quiescence of the HISCs. Furthermore, strong propidium iodide signals were detected in transformed ISCs from $esg^{ts} > N^{DN} + Arf79F^{RNAi}$ but not $esg^{ts} > N^{DN}$ flies (Extended data Fig. 6a–d), indicating that the transformed stem cells were dying through necrosis.

Similarly, DHE (Extended Data Fig. 6e–h) or LysoTracker (Extended Data Fig. 6i–l) signals were detected only in ISCs from $esg^{ts} > Arf79F^{RNAi}$ flies (Extended Data Fig. 6g, h, k, l), but not from wild-type flies (Extended Data Fig. 6e, f, i, j), indicating that the dying ISCs had accumulated ROS and were intracellularly acidified. Overexpressing *catalase* (a ROS-chelating enzyme) rescued the stem-cell death specifically induced by the *γ-COP* mutant clone (Fig. 2k) or by *Arf79F* knockdown (Extended Data Fig. 7b), and the ROS inhibitor NAC blocked the Arf1 inhibitor-induced death of Ras^{V12}-induced RNSC tumours (Extended Data Fig. 4i, l). These data together suggest that knockdown of the COPI–Arf1 complex induced the death of stem cells or of transformed stem cells (Ras^{V12}–RNSCs, N^{DN}–ISCs) through ROS-induced necrosis. Although ISCs, RNSCs, and HISCs exhibit different degrees of quiescence, they all rely on lipolysis for survival, suggesting that this is a general property of stem cells.

We noticed cases where the GFP-positive material of the dying ISCs was present within neighbouring enterocytes (Fig. 4a, Extended Data Fig. 5a–i), suggesting that these enterocytes had engulfed dying ISCs.

The JNK pathway, autophagy and engulfment genes are involved in the engulfment of dying cells[23,24]. We therefore investigated whether these genes are required for COPI–Arf79F-regulated

ISC death. We found that: (1) ISC death activated JNK signalling and autophagy in neighbouring enterocytes (Extended Data Fig. 7i–n); (2) knockdown of these genes in enterocytes but not in ISCs rescued ISC death to different degrees (Fig. 4b and Extended Data Figs 8a–i, 9a–l); (3) the drpr–mbc–Rac1–JNK pathway in enterocytes is not only necessary but also sufficient for ISC death (Extended Data Figs 8j–n and 9m, n); and (4) inhibitors of JNK and Rac1 could block Arf1-inhibitor-induced cell death of the Ras^{V12}-induced RNSC tumours (Extended Data Fig. 4g, h, l). These data together suggest that the drpr–mbc–Rac1–JNK pathway in neighbouring differentiated cells controls the engulfment of dying or transformed stem cells (Fig. 4c).

Our finding that the COPI–Arf79F–lipolysis–β-oxidation pathway regulated transformed stem-cell survival in the fly led us to investigate whether the pathway has a similar role in CSCs. We tested two Arf1 inhibitors (brefeldin A and golgicide A) and two FAO inhibitors (triascin C and etomoxir) on human cancer cell lines, and found that the growth, tumoursphere formation and expression of tumour-initiating cell markers of the four cancer cell lines were significantly ($P < 0.01$) suppressed by these inhibitors (Extended Data Fig. 10), suggesting that these inhibitors suppress CSCs. In mouse xenografts of BSY-1 human breast cancer cells, a novel low-cytotoxicity Arf1-ArfGEF inhibitor called AMF-26 was reported to induce complete regression *in vivo* in five days[25]. Together, this report and our results suggest that inhibiting Arf1 activity or blocking the lipolysis pathway can kill CSCs and block tumour growth.

Stem cells or CSCs are usually localized to a hypoxic storage niche, surrounded by a dense extracellular matrix[26], which may make them less accessible to sugar and amino acid nutrition from the body's circulatory system. Most normal cells rely on sugar and amino acids for their energy supply, with lipolysis playing only a minor role in their survival. Our results suggest that stem cells and CSCs are metabolically unique; they rely mainly on lipid reserves for their energy supply, and blocking COPI–Arf1-mediated lipolysis can starve them to death. We further found that transformed stem cells were more sensitive than normal stem cells to Arf1 inhibitors (Extended Data Fig. 4). Thus, selectively blocking lipolysis may kill CSCs without severe side effects. Therefore, targeting the COPI–Arf1 complex or the lipolysis pathway may prove to be a well-tolerated, novel approach for eliminating CSCs.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 29 July 2015; accepted 18 August 2016.
Published online 28 September 2016.

1. Kleffel, S. & Schatton, T. Tumor dormancy and cancer stem cells: two sides of the same coin? *Adv. Exp. Med. Biol.* **734,** 145–179 (2013).
2. Micchelli, C. A. & Perrimon, N. Evidence that stem cells reside in the adult *Drosophila* midgut epithelium. *Nature* **439,** 475–479 (2006).
3. Ohlstein, B. & Spradling, A. The adult *Drosophila* posterior midgut is maintained by pluripotent stem cells. *Nature* **439,** 470–474 (2006).
4. Singh, S. R., Liu, W. & Hou, S. X. The adult *Drosophila* malpighian tubules are maintained by multipotent stem cells. *Cell Stem Cell* **1,** 191–203 (2007).
5. Fox, D. T. & Spradling, A. C. The *Drosophila* hindgut lacks constitutively active adult stem cells but proliferates in response to tissue damage. *Cell Stem Cell* **5,** 290–297 (2009).
6. Takashima, S., Mkrtchyan, M., Younossi-Hartenstein, A., Merriam, J. R. & Hartenstein, V. The behaviour of *Drosophila* adult hindgut stem cells is controlled by Wnt and Hh signalling. *Nature* **454,** 651–655 (2008).
7. Essers, M. A. G. & Trumpp, A. Targeting leukemic stem cells by breaking their dormancy. *Mol. Oncol.* **4,** 443–450 (2010).
8. Jiang, H. *et al.* Cytokine/Jak/Stat signaling mediates regeneration and homeostasis in the *Drosophila* midgut. *Cell* **137,** 1343–1355 (2009).
9. Ma, M. *et al.* Wildtype adult stem cells, unlike tumor cells, are resistant to cellular damages in *Drosophila*. *Dev. Biol.* **411,** 207–216 (2016).
10. Zeng, X., Singh, S. R., Hou, D. & Hou, S. X. Tumor suppressors Sav/Scrib and oncogene Ras regulate stem-cell transformation in adult *Drosophila* malpighian tubules. *J. Cell. Physiol.* **224,** 766–774 (2010).
11. Ohlstein, B. & Spradling, A. Multipotent *Drosophila* intestinal stem cells specify daughter cell fates by differential notch signaling. *Science* **315,** 988–992 (2007).
12. Goulas, S., Conder, R. & Knoblich, J. A. The Par complex and integrins direct asymmetric cell division in adult intestinal stem cells. *Cell Stem Cell* **11,** 529–540 (2012).
13. Wu, M., Pastor-Pareja, J. C. & Xu, T. Interaction between Ras(V12) and scribbled clones induces tumour growth and invasion. *Nature* **463,** 545–548 (2010).
14. Lee, M. C. S., Miller, E. A., Goldberg, J., Orci, L. & Schekman, R. Bi-directional protein transport between the ER and Golgi. *Annu. Rev. Cell Dev. Biol.* **20,** 87–123 (2004).
15. Beller, M. *et al.* COPI complex is a regulator of lipid homeostasis. *PLoS Biol.* **6,** e292 (2008).
16. Zeng, X. *et al.* Genome-wide RNAi screen identifies networks involved in intestinal stem cell regulation in *Drosophila*. *Cell Reports* **10,** 1226–1238 (2015).
17. Lee, T. & Luo, L. Mosaic analysis with a repressible cell marker for studies of gene function in neuronal morphogenesis. *Neuron* **22,** 451–461 (1999).
18. Zhang, Y., Chen, D. & Wang, Z. Analyses of mental dysfunction-related ACSl4 in *Drosophila* reveal its requirement for Dpp/BMP production and visual wiring in the brain. *Hum. Mol. Genet.* **18,** 3894–3905 (2009).
19. Min, K. T. & Benzer, S. Preventing neurodegeneration in the *Drosophila* mutant bubblegum. *Science* **284,** 1985–1988 (1999).
20. Grönke, S. *et al.* Brummer lipase is an evolutionary conserved fat storage regulator in *Drosophila*. *Cell Metab.* **1,** 323–330 (2005).
21. Palanker, L., Tennessen, J. M., Lam, G. & Thummel, C. S. *Drosophila* HNF4 regulates lipid mobilization and β-oxidation. *Cell Metab.* **9,** 228–239 (2009).
22. Timmons, A. K., Meehan, T. L., Gartmond, T. D. & McCall, K. Use of necrotic markers in the *Drosophila* ovary. *Methods Mol. Biol.* **1004,** 215–228 (2013).
23. Etchegaray, J. I. *et al.* Draper acts through the JNK pathway to control synchronous engulfment of dying germline cells by follicular epithelial cells. *Development* **139,** 4029–4039 (2012).
24. Ohsawa, S. *et al.* Elimination of oncogenic neighbors by JNK-mediated engulfment in *Drosophila*. *Dev. Cell* **20,** 315–328 (2011).
25. Ohashi, Y. *et al.* AMF-26, a novel inhibitor of the Golgi system, targeting ADP-ribosylation factor 1 (Arf1) with potential for cancer therapy. *J. Biol. Chem.* **287,** 3885–3897 (2012).
26. Trumpp, A. & Wiestler, O. D. Mechanisms of disease: cancer stem cells—targeting the evil twin. *Nat. Clin. Pract. Oncol.* **5,** 337–347 (2008).
27. Han, C. Z. & Ravichandran, K. S. Metabolic connections during apoptotic cell engulfment. *Cell* **147,** 1442–1445 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank S. Hayashi, J.-P. Vincent, M. Fortini, C. Thummel, E. Baehrecke, R. P. Kuhnlein, M. Freeman, F. Schweisguth, M. Mlodzik, T. Lecuit, DGRC, VDRC, and the Bloomington Stock Centers for fly stocks; A. Chavanieu and L. Frigerio for Arf1 inhibitors; X. Yang and the Developmental Studies Hybridoma Bank for antibodies; and S. Lockett for help with the confocal microscope. This research was supported by the Intramural Research Program of the National Institutes of Health, National Cancer Institute.

Author Contributions S.X.H., X.Z. and S.R.S. conceived and designed the experiments. S.R.S., X.Z. and S.X.H. performed the *Drosophila* experiments. J.Z. performed the experiments using human cancer cell lines. Y.L., G.H. and H.L. assisted with experiments. S.X.H., S.R.S., X.Z. and J.Z. analysed the data. S.X.H., S.R.S. and J.Z. wrote the manuscript. S.X.H. and S.R.S. revised the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.X.H. (hous@mail.nih.gov).

Reviewer Information *Nature* thanks Y. Apidianakis, M. Montminy, H. Steller and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Fly strains. The following fly strains were used: *NP1–Gal4* and *FRT19A–δ-COP*^{G0051} (from DGRC); *esg–Gal4* and *tsh–Gal4* (from S. Hayashi); *wg–Gal4* (from J.-P. Vincent); *UAS–upd* (generated in our laboratory); *UAS–N*$^{DN(2X)}$ (from M. Fortini); *UAS–bmm* (from R. P. Kuhnlein); *UAS–Hnf4* and *GAL4–dHFN4 ; UAS–nlacZ* (from C. Thummel); *mira–GFP* (from F. Schweisguth); *pmCherry–Atg8a* (from E. Baehrecke); *UAS–GAP43–mCherry* (from T. Lecuit); *UAS–drpr*^{RNAi} and *UAS–drpr* (from M. Freeman); *UAS–DJun*^{Asp} (from M. Mlodzik); *UAS–hep*^{CA}, *UAS–Ras*^{V12}, *UAS–scu*, *UAS–rpr*, *UAS–p53*, *UAS–p35*, *UAS–Cat*, *UAS–Sod*, *UAS–Sod2*, *UAS–Rac*$^{DN(N17)}$, *UAS–Rac*^{V12}, *UAS–bsk*^{DN}, *hop*$^{Tum-l}$, *puc–lacZ* (*puc*^{E69}), *UAS–2XEYFP*, *tub–Gal80*^{ts}, and fly lines used for MARCM clones (*FRT*^{19A}, *tub–Gal80*; *FRT*^{82B}, *tub–Gal80*; *SM6, hs–flp*; *MKRS, hs–flp*; *act>y+>Gal4, UAS–GFP*; *FRT*^{82B}*–γ-COP*^{10}; and UAS–*γ-COP FRT*^{82B}*–γ-COP*^{10}) were obtained from the Bloomington *Drosophila* Stock Center at Indiana University.

RNAi stocks. An upstream activating sequence (UAS)-regulated double-stranded inverse-repeat construct was designed to target *Arf79F*: (*UAS–Arf79F*^{RNAi})-VDRC Transformant ID: 23082 (v23082). The RNA level was reduced to 39.0% in the *Act–Gal4–UAS–Arf79F*^{RNAi} flies (ref. 16), and the phenotypes were confirmed by two independent RNAi lines (v103572 and v23080). The other RNAi lines used were: *δ-COP*–v41551; RNA level was reduced to 14.3% in *Act–Gal4–UAS–δ-COP*^{RNAi} flies (ref. 16), and phenotypes were confirmed by an independent RNAi line (Bloomington stock number 31764 (BL31764 (TRiP ID HM04076)). *β-COP*–BL31709 (TRiP ID HM04016); RNA level was reduced to 13.3% in *Act–Gal4–UAS–β-COP*^{RNAi} flies (ref. 16), and phenotypes were confirmed by two independent RNAi lines (v109641 and v15418). *β'-COP*–BL31710 (TRiP ID HM04017); RNA level was reduced to 3.2% in *Act–Gal4–UAS–β'-COP*^{RNAi} flies (ref. 16). *ζ-COP*-BL28960 (TRiP ID HM05171); RNA level was reduced to 47.0% in *Act–Gal4–UAS–ζ-COP*^{RNAi} flies (ref. 16), and phenotypes were confirmed by two independent RNAi lines (v34768 and v104405). *garz*-BL31232 (TRiP ID JF01013); RNA level was reduced to 52.4% in the *Act–Gal4–UAS–garz*^{RNAi} flies (ref. 16). *Acsl*–BL27729 (TRiP ID JF02811); RNA level was reduced to 25.5% in *Act–Gal4–UAS–Acsl*^{RNAi} flies (ref. 16). *bgm*-v34854; RNA level was reduced to 56.2% in *Act–Gal4–UAS–bgm*^{RNAi} flies (ref. 16), and phenotypes were confirmed by two independent RNAi lines (v105635 and BL28639 (TRiP ID JF03054)). *γ-COP*–BL28889 (TRiP IDHM05099); *Atg5*–BL34899 (TRiP ID HMS01244); *Atg12*^{RNAi} (from E. Baehrecke, ref. 28); *mbc*–BL32355 (TRiP ID HMS00346); *PSR*–BL33700 (TRiP ID HMS00576); *mys*–BL33642 (TRiP ID HMS00043); *CycT*–BL32976 (TRiP ID HMS00776). The sequences used for each VDRC knock-down strain are available at https://stockcenter.vdrc.at) and for each Bloomington knock-down strain at http://flystocks.bio.indiana.edu. The data presented in all figures were generated by using the first RNAi line for all genes.

MARCM clone assay. To induce MARCM clones, three- or four-day-old adult female flies were heat-shocked twice with an interval of 8–12 h at 37 °C for 60 min. The flies were transferred to fresh food daily after the final heat shock and their posterior midguts were processed for staining at the indicated times.

RNAi-mediated gene depletion. To target the expression of UAS-linked genes in the cell types of interest, we used specific drivers. The posterior midgut of adult *Drosophila* is maintained by multipotent ISCs[2,3], which differentiate into secretory enteroendocrine cells and absorptive enterocytes through immature enteroblasts. Enterocytes are polyploid and express the transcription factor Pdm1. Enteroendocrine cells are diploid and express the transcription factor Prospero (Pros). UAS-linked genes can be targeted to enterocytes by the *MyolAGal4* (*NP1–Gal4*) driver[8] or to ISCs and enteroblasts by the *escargot* (*esg*)*–Gal4* driver[2]. To target expression of UAS-linked genes in RNSCs, we also used *esg–Gal4* (ref. 4); for the quiescent HISCs, the *wingless* (*wg*)*–Gal4* driver was used[5,6]. To investigate the response of the different cells to cell-death effectors, we first overexpressed *reaper* (*rpr*, an inhibitor of *Death-associated inhibitor of apoptosis 1*; *Diap-1*) in them, using the cell-type-specific Gal4 drivers combined with the temperature-sensitive Gal4 repressor *tub–Gal80*^{ts} (ref. 29).

We used the inducible *NP1–Gal4; tub–Gal80*^{ts}*–UAS–rpr; UAS–GFP* to express *rpr* in enterocytes (*NP1*^{ts} > *rpr*), *esg–Gal4; tubGal80*^{ts}*–UAS–rpr; UAS–GFP* (*esg*^{ts} > *rpr*) in ISCs, enteroblasts and RNSCs, and *wg–Gal4; tubGal80*^{ts}*–UAS–rpr; UAS–GFP* (*wg*^{ts} > *rpr*) in HISCs. The *NP1*^{ts} > *rpr* flies were raised to adulthood at 18 °C and shifted to 29 °C for 24 h to induce *rpr* expression.

Four male *UAS–RNAi* transgene flies were crossed with 8 female virgins of *NP1*^{ts}, *esg*^{ts}, and *wg*^{ts} at 18 °C. Three- to five-day-old adult flies with the appropriate genotype were transferred to new vials at 29 °C for the indicated times before dissection. For p53, we did not find a significant change in *esg*$^{+}$ progenitors and enteroendocrine cell numbers after the flies (*esg*^{ts} > *p53*) were cultured for 7 days at 29 °C, although a previous study found that a 15-day induction ablated nearly all *esg*$^{+}$ cells and reduced enteroendocrine cell numbers[8].

Histology and image capture. Fly intestines were dissected in PBS and fixed in PBS containing 4% formaldehyde for 30 min. After three 5-min rinses with 1× PBT (PBS + 0.1% Triton X-100), the samples were blocked in 1× PBT containing 5% normal goat serum overnight at 4 °C and incubated first with primary antibody overnight at 4 °C or at room temperature for 2 h, and then with a fluorescence-conjugated secondary antibody for 2 h at room temperature. Samples were mounted in Vectashield mounting medium with DAPI (Vector Laboratories). The following antibodies were used: rabbit polyclonal anti-β-Gal (1:1,000; Cappel); mouse anti-Dl (Delta, 1:20; DSHB); mouse monoclonal anti-Prospero (Pros, 1:50; DSHB); rabbit polyclonal anti-Pdm1 (1:1,000, a gift from X. Yang); mouse monoclonal anti-Arm N27A1 (1:20; DSHB); Rabbit monoclonal anti-Phospho-SAPK/JNK (1:200; Cell Signaling); rabbit-polyclonal anti-GFP (1:500, Invitrogen); mouse monoclonal anti-GFP (1:100; Invitrogen), and chicken polyclonal anti-GFP (1:3,000; Abcam). Secondary antibodies were goat anti-mouse, anti-chicken, and goat anti-rabbit IgG conjugated to Alexa488 or Alexa568 (1:400; Invitrogen). DAPI (Sigma) was used to stain DNA. CellMask Deep Red plasma membrane dye (Life Technologies, C10046) was used to visualize the plasma membrane. Midguts were labelled with CellMask Deep Red Plasma Membrane Stains (1:2,000) for 7 min[30].

ROS detection by DHE. DHE staining was performed as described previously[31]. In brief, guts were dissected in 1× PBS, incubated in 30 μM DHE (Invitrogen) in PBS for 5 min at room temperature in the dark, washed twice, mounted and immediately imaged by confocal microscopy.

Lysotracker staining. Guts were dissected in 1× PBS and then stained without fixation in 0.5 μM Lysotracker Red DND-99 (Invitrogen) for 3 min at room temperature. They were then washed three times in 1× PBS, fixed for 20 min in 4% formaldehyde, washed three times in 1× PBT, rinsed twice with 1× PBS, mounted in Vectashield with DAPI and analysed on a confocal microscope.

Apoptosis detection. Apoptosis was detected by TUNEL with the ApopTag Red *In situ* Apoptosis Detection Kit (Chemicon International) according to the manufacturer's instructions.

Necrosis evaluation by propidium iodide. Guts were dissected in 1× PBS and then stained in 1.5 μM PI (Invitrogen) for 15 min at room temperature. The guts were then fixed for 20 min in 4% formaldehyde, washed three times in 1× PBT, rinsed twice with 1× PBS, mounted in Vectashield with DAPI and analysed on a confocal microscope.

Oil Red O staining. Oil Red O staining was performed as described previously[32]. In brief, *Drosophila* midguts were dissected in 1× PBS and fixed in 4% formaldehyde for 30 min. Midguts were washed three times in 1× PBS, double-distilled water and a 60% isopropanol solution. From the stock solution of Oil Red O (Sigma-Aldrich; 0.1% solution in isopropanol), a working solution was prepared by mixing 6 ml of 0.1% Oil Red O in isopropanol and 4 ml of double-distilled water. Midguts were incubated for 20 min in this solution and then washed in 60% isopropanol and water. The midguts were mounted in Vectashield mounting medium with DAPI (Vector Laboratories) and were imaged by confocal microscopy. Images were captured with the Zeiss LSM 510 confocal system and processed with LSM Image Browser and Adobe Photoshop.

Quantification and statistical analysis. To determine the percentage of GFP^{+} cells, the GFP^{+} cells and total cells were counted in a 5,000-μm^{2} area of a single confocal plane. In *esg*^{ts} samples, cells were counted in the posterior midgut and Malpighian tubules; in *wg*^{ts} samples, they were counted in the hindgut–midgut junction; and in *esg*^{ts}*wg*^{ts} samples, they were counted in the hindgut–midgut junction, the posterior midgut and the Malpighian tubules. The number of Pros^{+} nuclei was counted in a 0.08-mm^{2} surface area of a microscopic image from a similar region of each posterior midgut[33]. Cells per tumour were determined by counting the total number of nuclei within GFP^{+} tumours. All of the images were taken with the LSM5 Image Browser using the same confocal settings (Zeiss). Statistical analyses were performed using GraphPad Prism. Sample sizes (n) reported reflect the number of individual midguts. All experiments were performed in triplicate. P values were obtained between two groups using the Student's t-test. For all statistical analysis, differences were considered to be statistically significant at values of $P < 0.05$.

Flow cytometry. Cell surface markers were analysed by flow cytometry. Cultured or treated cells were dissociated by 0.05% trypsin-EDTA and centrifuged. Single cells were resuspended with PBS containing 2% FBS and fluorescent-conjugated antibodies FITC-CD44 (clone G44-26, BD Biosciences) and PE-CD24 (clone ML5, Biolegend), and incubated on ice for 30 min. After washing three times with PBS containing 2% FBS, cells were resuspended with PBS containing 2% FBS and analysed by BD FACS Caliber (BD Biosciences).

Inhibitors. Arf1 inhibitors are BFA (brefeldin A)[34] from Sigma, GCA (golgicide A)[35] from Santa Cruz, secin H3[36] from Cayman chemical, LM11[37] from A. Chavanieu, LG8[38] from L. Frigerio. 2-deoxy-D-Glucose (2-DG)[39]; JNK inhibitor SP 600125[40]; FAO inhibitors:triascin C[41], etomoxir[42], and mildronate[43] from

Cayman chemical, Enoximone[44] from Tocris. Rac1 inhibitor from Santa Cruz. N-Acetyl-L-cysteine (NAC) [45] from Sigma. For control experiments, a DMSO control (100 μl in 10 ml food) was used. All inhibitors were mixed in the fly food with following concentrations; Arf1 inhibitors: BFA (50 ng ml^{-1} and 200 ng ml^{-1}), GCA (5 μM), LM11 (50 μM), LG8 (100 μM), and secin H3 (50 μM); JNK inhibitor: Sp600125 (50 μM); Rac1 inhibitor (100 μM); ROS inhibitor: NAC (10 mM); FAO inhibitors: triacsin C (5 μM), mildronate (100 μM), etomoxir (100 μM), enoximone (100 μM); and glycolysis inhibitor 2-DG (50 mM). We mixed each inhibitor in fly food and tested different concentrations of these inhibitors. We used the concentration in which the inhibitors could kill tumour cells. At the beginning of experiments, to find out whether the flies would eat the inhibitors, we added green food dye to fly food. Flies were fasted for 1 h and then 8–10 flies were transferred to a vial containing coloured fly food mixed with inhibitors. We found that within 1 h of feeding the green dye could be seen through the abdomen of each fly, which suggest that the fly food mixed with inhibitors was edible to the flies. However, we excluded the food dye from the food used in the main experiments. For the main experiments, we fed flies with food containing inhibitors for 4 days. We repeated each inhibitor treatment three times.

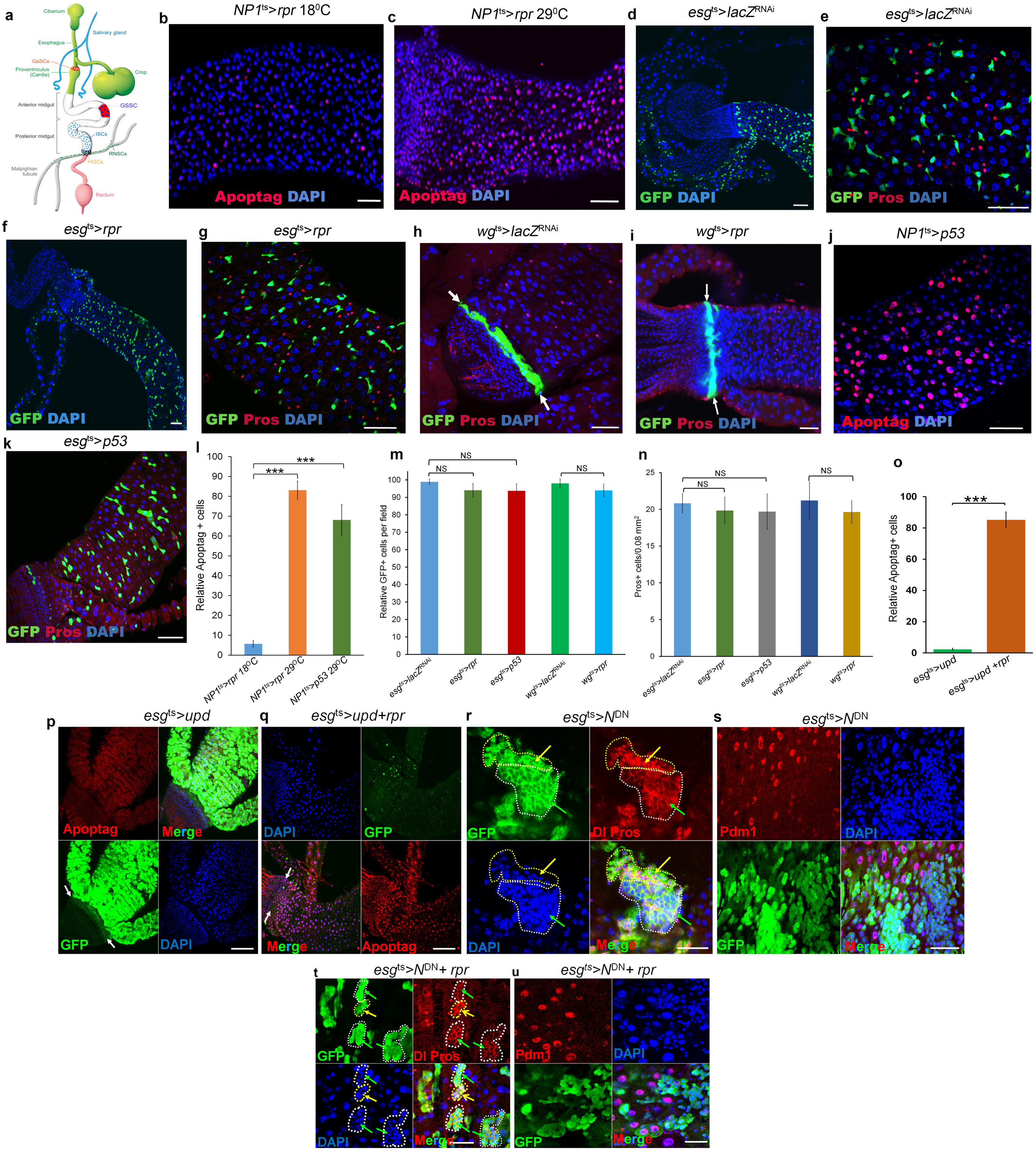
Cell lines and culture. The human prostate cancer cell line DU145, colon cancer cell line HT29, and breast cancer cell lines MCF7, MDA-MB-231 (provided by the DCTD Tumour Repository) were cultured in RPMI1640 supplemented with 10% fetal bovine serum and 100 units per ml penicillin/streptomycin at 37 °C in a humidified atmosphere containing 5% CO_2.

Cell survival assay. Cells were seeded at 2×10^5 cells per well in 6-well plates. Treatments with the indicated chemicals were started the next day, and cells were incubated for 2 more days. The surviving cells were stained with Crystal Violet (EMD Millipore). For quantification, the stained cells were solubilized in 1% SDS, and the absorbance at 595 nm was determined with a microplate reader.

Sphere formation assay. Single cells were cultured in a Corning Costar Ultra-Low attachment 24-well plate (Sigma-Aldrich) in sphere culture medium, consisting of DMEM/F12 (1:1), B27 (Invitrogen), and 20 ng ml^{-1} EGF (Invitrogen), with the indicated chemicals. The number of spheres was counted after 10 days of culture.

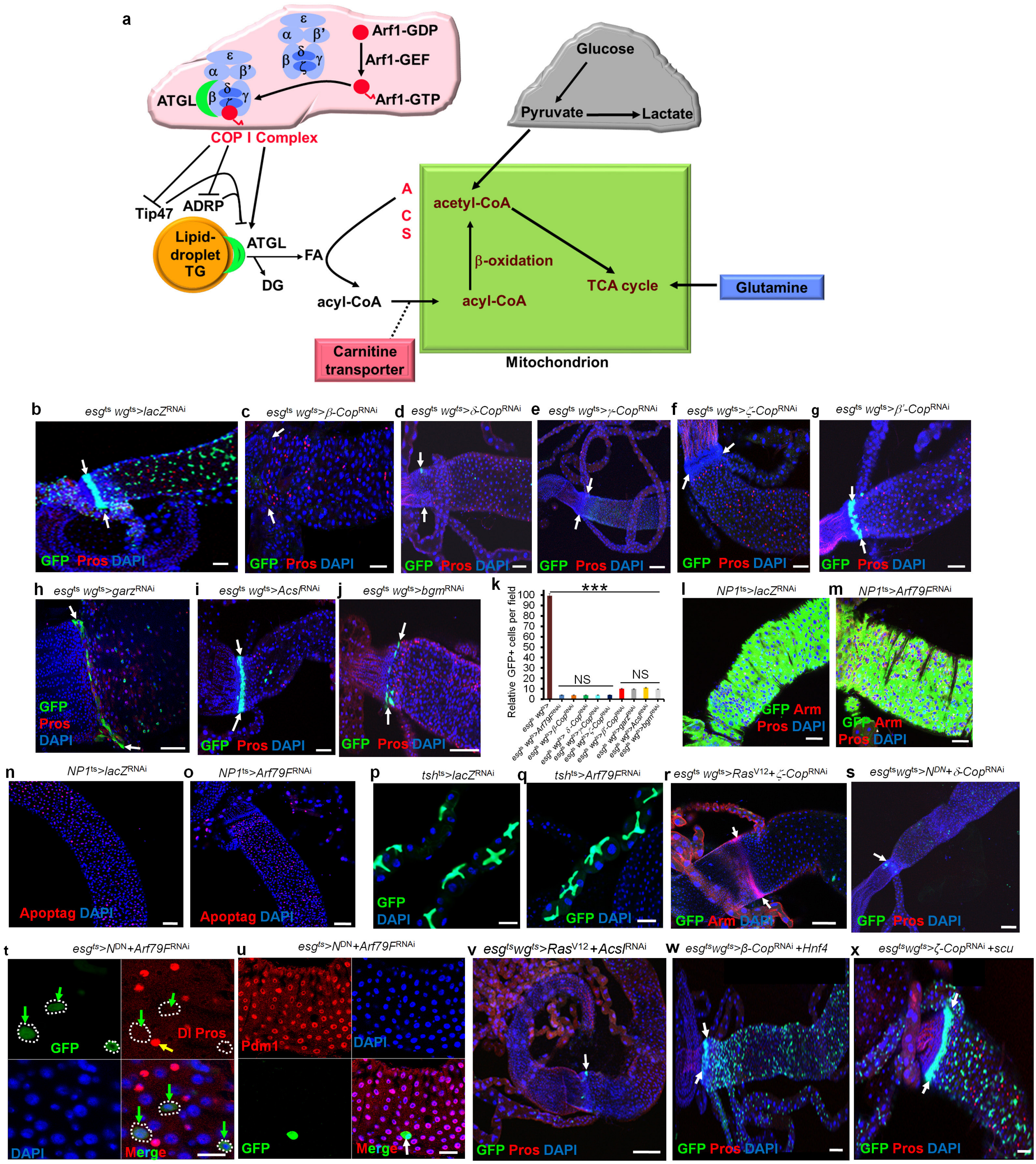
RNA isolation and real-time PCR. An RNeasy Mini Kit (Qiagen) was used to extract the total RNA from human cancer cells. The cDNA was synthesized from 1 μg RNA from each sample using a reverse transcription kit (Promega). Real-time PCR was performed in a 15-μl reaction system using SYBR Advantage qPCR Premix (Clontech). All of the reactions were performed in triplicate in a RealPlex 2 system (Eppendorf). The relative gene expression was quantified as described previously[46]. The sequence of each primer was as follows: *ACTB*, 5′-GATCATTGCTCCTCCTGAGC-3′ and 5′-ACTCCTGCTTGCTGATCCAC-3′;*CDH1*, 5′-ACCAGAATAAAGACCAAGTGACCA-3′ and 5′-AGCAAGAGCAGCAGAATCAGAAT-3′; *CD44*, 5′-GAGCATCGGATTTGAGA-3′ and 5′-CATACTGGGAGGTGTTGG-3′.

28. McPhee, C. K., Logan, M. A., Freeman, M. R. & Baehrecke, E. H. Activation of autophagy during cell death requires the engulfment receptor Draper. *Nature* **465,** 1093–1096 (2010).
29. McGuire, S. E., Le, P. T., Osborn, A. J., Matsumoto, K. & Davis, R. L. Spatiotemporal rescue of memory dysfunction in *Drosophila*. *Science* **302,** 1765–1768 (2003).
30. Hsieh, H. H., Chang, W. T., Yu, L. & Rao, Y. Control of axon-axon attraction by Semaphorin reverse signaling. *Proc. Natl. Acad. Sci. USA* **111,** 11383–11388 (2014).
31. Hochmuth, C. E., Biteau, B., Bohmann, D. & Jasper, H. Redox regulation by Keap1 and Nrf2 controls intestinal stem cell proliferation in *Drosophila*. *Cell Stem Cell* **8,** 188–199 (2011).
32. Reiff, T. *et al.* Endocrine remodelling of the adult intestine sustains reproduction in *Drosophila. eLife* **4,** e06930 (2015).
33. Amcheslavsky, A. *et al.* Enteroendocrine cells support intestinal stem-cell-mediated homeostasis in *Drosophila*. *Cell Reports* **9,** 32–39 (2014).
34. Sausville, E. A.*et al.* Antiproliferative effect in vitro and antitumor activity in vivo of brefeldin A. *Cancer J. Sci. Am.* **2,** 52–58 (1996).
35. Sáenz, J. B. *et al.* Golgicide A reveals essential roles for GBF1 in Golgi assembly and function. *Nat. Chem. Biol.* **5,** 157–165 (2009).
36. Hafner, M. *et al.* Inhibition of cytohesins by SecinH3 leads to hepatic insulin resistance. *Nature* **444,** 941–944 (2006).
37. Viaud, J. *et al.* Structure-based discovery of an inhibitor of Arf activation by Sec7 domains through targeting of protein-protein complexes. *Proc. Natl Acad. Sci. USA* **104,** 10370–10375 (2007).
38. Sorieul, M. *et al.* An Exo2 derivative affects ER and Golgi morphology and vacuolar sorting in a tissue-specific manner in *arabidopsis*. *Traffic* **12,** 1552–1562 (2011).
39. Ciavardelli, D. *et al.* Breast cancer stem cells rely on fermentative glycolysis and are sensitive to 2-deoxyglucose treatment. *Cell Death Dis.* **5,** e1336 (2014).
40. Bennett, B. L. *et al.* SP600125, an anthrapyrazolone inhibitor of Jun N-terminal kinase. *Proc. Natl Acad. Sci. USA* **98,** 13681–13686 (2001).
41. Mashima, T. *et al.* p53-defective tumors with a functional apoptosome-mediated pathway: a new therapeutic target. *J. Natl. Cancer Inst.* **97,** 765–777 (2005).
42. Samudio, I. *et al.* Pharmacologic inhibition of fatty acid oxidation sensitizes human leukemia cells to apoptosis induction. *J. Clin. Invest.* **120,** 142–156 (2010).
43. Dambrova, M., Liepinsh, E. & Kalvinsh, I. Mildronate: cardioprotective action through carnitine-lowering effect. *Trends Cardiovasc. Med.* **12,** 275–279 (2002).
44. Abdel-Aleem, S., Youssef, J., Frangakis, C. & Badr, M. Selective inhibition of hepatic peroxisomal fatty acid beta-oxidation by enoximone. *Life Sci*. **51,** 53–57 (1992).
45. Delneste, Y., Jeannin, P., Potier, L., Romero, P. & Bonnefoy, J. Y. N-acetyl-L-cysteine exhibits antitumoral activity by increasing tumor necrosis factor alpha-dependent T-cell cytotoxicity. *Blood* **90,** 1124–1132 (1997).
46. Feng, X. *et al.* Selection of reference genes for gene expression studies in porcine skeletal muscle using SYBR green qPCR. *J. Biotechnol.* **150,** 288–293 (2010).
47. Singh, S. R., Zeng, X., Zheng, Z. & Hou, S. X. The adult *Drosophila* gastric and stomach organs are maintained by a multipotent stem cell pool at the foregut/midgut junction in the cardia (proventriculus). *Cell Cycle* **10,** 1109–1120 (2011).
48. Strand, M. & Micchelli, C. A. Quiescent gastric stem cells maintain the adult *Drosophila* stomach. *Proc. Natl Acad. Sci. USA* **108,** 17696–17701 (2011).
49. O'Brien, L. E., Soliman, S. S., Li, X. & Bilder, D. Altered modes of stem cell division drive adaptive intestinal growth. *Cell* **147,** 603–614 (2011).
50. Kirilly, D., Spana, E. P., Perrimon, N., Padgett, R. W. & Xie, T. BMP signaling is required for controlling somatic stem cell self-renewal in the *Drosophila* ovary. *Dev. Cell* **9,** 651–662 (2005).
51. Adachi-Yamada, T., Fujimura-Kamada, K., Nishida, Y. & Matsumoto, K. Distortion of proximodistal information causes JNK-dependent apoptosis in *Drosophila* wing. *Nature* **400,** 166–169 (1999).
52. Zhou, Z., Hartwieg, E. & Horvitz, H. R. CED-1 is a transmembrane receptor that mediates cell corpse engulfment in *C. elegans*. *Cell* **104,** 43–56 (2001).
53. Ziegenfuss, J. S., Doherty, J. & Freeman, M. R. Distinct molecular pathways mediate glial activation and engulfment of axonal debris after axotomy. *Nat. Neurosci.* **15,** 979–987 (2012).
54. Fadok, V. A. *et al.* A receptor for phosphatidylserine-specific clearance of apoptotic cells. *Nature* **405,** 85–90 (2000).
55. Wu, Y., Tibrewal, N. & Birge, R. B. Phosphatidylserine recognition by phagocytes: a view to a kill. *Trends Cell Biol*. **16,** 189–197 (2006).
56. Dontu, G. *et al.* In vitro propagation and transcriptional profiling of human mammary stem/progenitor cells. *Genes Dev.* **17,** 1253–1270 (2003).



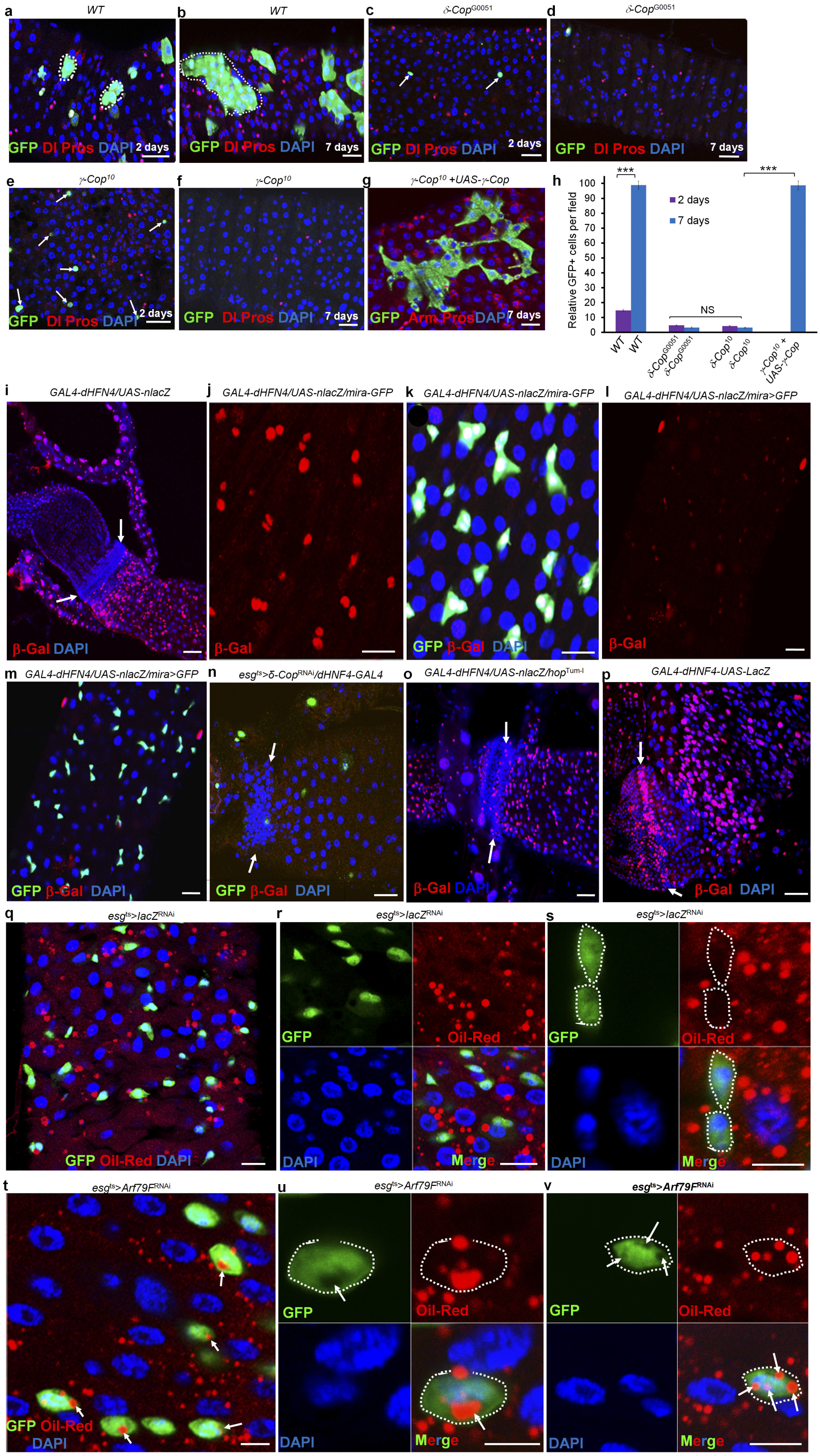
Extended Data Figure 1 | See next page for caption.

Extended Data Figure 1 | Stem cells are resistant to apoptosis. a, Stem cells in the adult *Drosophila* digestive system. In this system, three organs, the posterior midgut, the hindgut and the Malpighian tubules, meet and join at the junction of the posterior midgut and hindgut. Stem cells in these organs exhibit different degrees of quiescence. The intestinal stem cells (ISCs), located in the posterior midgut, divide once every 24 h[2,3]; the renal and nephric stem cells (RNSCs), located in the Malpighian tubules, divide about once a week[4]; and the quiescent hindgut intestinal stem cells (HISCs), found at the midgut–hindgut junction, divide only during stress-induced tissue repair[5,6]. GaSCs are gastric stem cells at the foregut–midgut junction[47]. GSSCs are gastric stem cells in the middle of the midgut[48]. The colours just make the cell types or organs more visible and do not exactly reflect different regions in the digestive system. **b–n** Stem cells are resistant to apoptosis. **b**, $NP1^{ts} > rpr$, 18 °C, 24 h ($n = 37$). **c**, $NP1^{ts} > rpr$, 29 °C, 24 h ($n = 29$). **d**, **e**, $esg^{ts} > lacZ^{RNAi}$, 29 °C, 7 d ($n = 32$). **f**, **g**, $esg^{ts} > rpr$, 29 °C, 7 d ($n = 35$). **h**, $wg^{ts} > lacZ^{RNAi}$, 29 °C, 7 d ($n = 27$). **i**, $wg^{ts} > rpr$, 29 °C, 7 d ($n = 24$). **j**, $NP1^{ts} > p53$, 29 °C, 5 d ($n = 31$). **k**, $esg^{ts} > p53$, 29 °C, 7 d ($n = 38$). **l**, Quantification of $Apoptag^{+}$ cells in the indicated panels. **m**, Quantification of GFP^{+} cells in the indicated panels. **n**, Quantification of $Pros^{+}$ cells in the indicated panels. Data are represented as mean ± s.d.. Statistical significance determined by Student's *t*-test, $^{***}P < 0.0001$. NS, not significant ($P > 0.05$). As reported previously[8], 24-h induction of *rpr* in enterocytes resulted in widespread apoptosis (compare **c** with **b** and see the quantitative comparison in **l**). The induction of *rpr* by *esg-Gal4* (**f**, **g**) or *wg-Gal4* (**i**) had little effect on the progenitor or stem cells (that is, enteroblasts, ISCs, RNSCs and HISCs) at one week, compared to wild-type controls (compare **f**, **g** with **d**, **e**; **i** with **h**, and see the quantitative comparison in **m**). We also found that the overexpression of *Drosophila* p53 could effectively ablate the enterocytes in five days (compare **j** with **b** and see the quantitative comparison in **l**) but had little effect on stem cells even after one week, compared to controls (compare **k** with **e** and see the quantitative comparison in **m**). Because *NP1–Gal4* and *esg–Gal4* are not expressed in enteroendocrine cells, as expected, we did not find significant changes in enteroendocrine cells in these experiments (**n**). **o**–**u**, Activation of proliferation accelerates apoptotic cell death of hyperplastic stem cells but fails to completely eliminate neoplastic stem cells. **o**, Quantification of $Apoptag^{+}$ cells in the indicated panels. Data are represented as mean ± s.d. Statistical significance was determined by Student's *t*-test, $^{***}P < 0.0001$. **p**, $esg^{ts} > upd$, 29 °C, 4 d ($n = 28$). **q**, $esg^{ts} > upd + rpr$, 29 °C, 4 d ($n = 33$). **r**, **s**, $esg^{ts} > N^{DN}$, 29 °C, 7 d ($n = 25$). **t**, **u**, $esg^{ts} > N^{DN} + rpr$, 29 °C, 7 d ($n = 32$). White arrows point to the hindgut–midgut junction in **h**, **i**, **p**, **q**; yellow arrows point to $Pros^{+}$ enteroendocrine cells in **r** and **t**; green arrows point to Dl^{+} ISCs in **r** and **t**. White dotted lines outline GFP^{+} stem cell clusters in **r** and **t**. Yellow dotted lines outline enteroendocrine cell clusters in **r** and **t**. Expression of *rpr* or $Arf79F^{RNAi}$ in ISCs did not kill differentiated cells. The posterior midguts of flies with the indicated genotypes were dissected, stained with the indicated antibodies and analysed by confocal microscopy. Scale bars in **b**–**k** and **p**–**u**, 10 μm.



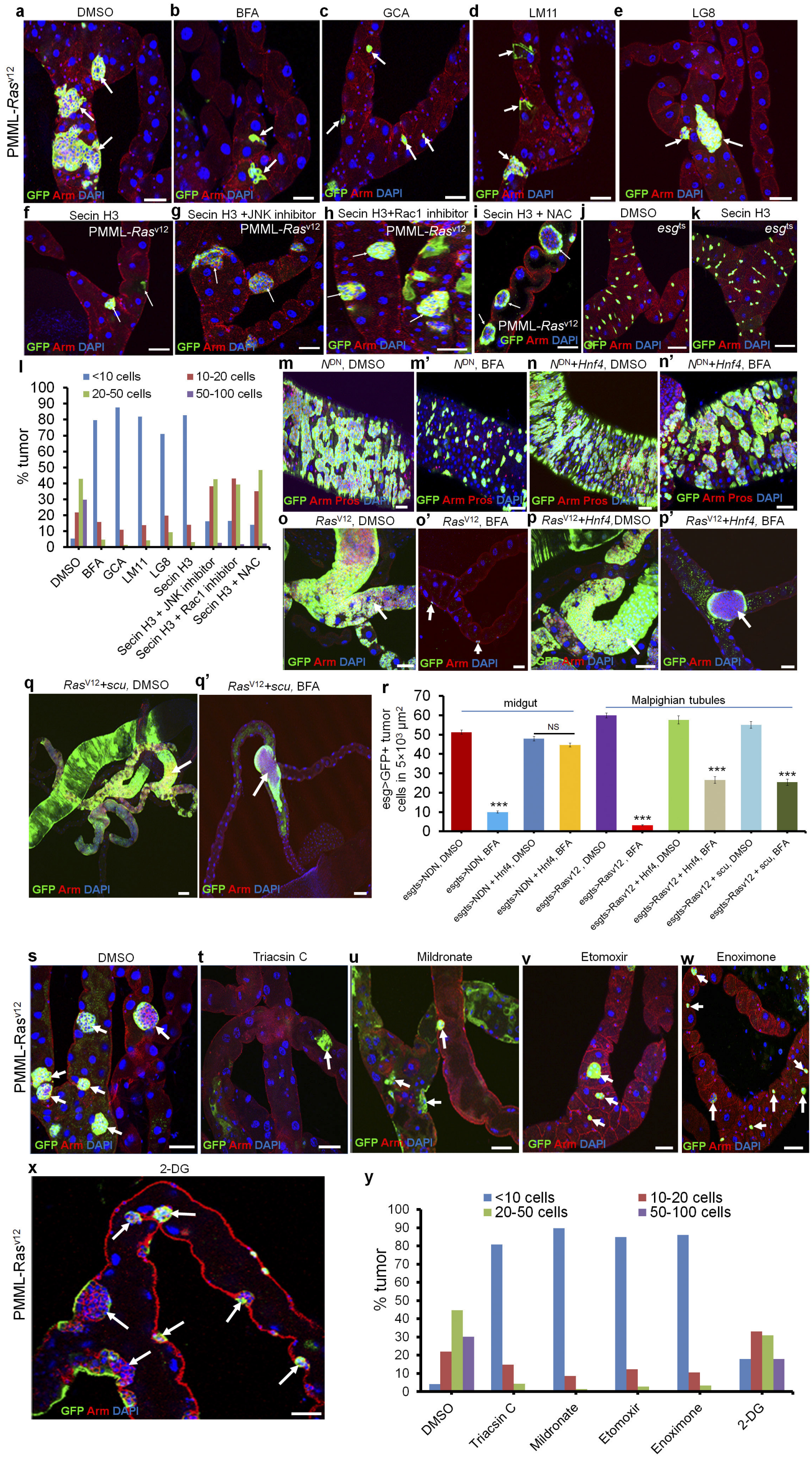
Extended Data Figure 2 | See next page for caption.

Extended Data Figure 2 | The COPI–Arf1 complex regulates stem but not differentiated cell survival. a, The COPI–lipolysis–β-oxidation pathway. The COPI–Arf1 complex controls lipid homeostasis by regulating adipocyte differentiation-related protein (ADRP), tail-interacting protein of 47kDa (Tip47) and adipocyte triglyceride lipase (ATGL)[15]. Triglycerides (TG), diglyceride (DG), fatty acid (FA), Acyl-CoA synthetase (ACS). **b**–**j**, The COPI–Arf1 complex regulates stem cell survival. The genotypes of the flies in each panel were: **b**, esg^{ts} $wg^{ts} > lacZ^{RNAi}$, 29 °C, 7 d (n = 38). **c**, esg^{ts} $wg^{ts} > \beta\text{-}Cop^{RNAi}$, 29 °C, 7 d ($n$ = 23). **d**, esg^{ts} $wg^{ts} > \delta\text{-}Cop^{RNAi}$, 29 °C, 7 d ($n$ = 32). **e**, esg^{ts} $wg^{ts} > \gamma\text{-}Cop^{RNAi}$, 29 °C, 7 d ($n$ = 27). **f**, esg^{ts} $wg^{ts} > \zeta\text{-}Cop^{RNAi}$, 29 °C, 7 d ($n$ = 31). **g**, esg^{ts} $wg^{ts} > \beta'\text{-}Cop^{RNAi}$, 29 °C, 7 d ($n$ = 29). **h**, esg^{ts} $wg^{ts} > garz^{RNAi}$, 29 °C, 7 d (n = 27). **i**, esg^{ts} $wg^{ts} > Acsl^{RNAi}$, 29 °C, 7 d (n = 32). **j**, esg^{ts} $wg^{ts} > bgm^{RNAi}$, 29 °C, 7 d (n = 25). **k**, Quantification of GFP^{+} cells in the indicated panels. Data show mean ± s.e.m. Statistical significance was determined by Student's t-test, $^{***}P < 0.0001$. NS, not significant ($P > 0.05$). **l**–**q** Knockdown of the COPI–Arf79F complex did not kill differentiated cells. The genotypes of the flies in each panel were: **l**, **n**, $NP1^{ts} > lacZ^{RNAi}$, 29 °C, 7 d (n = 25). **m**, **o**, $NP1^{ts} > Arf79F^{RNAi}$, 29 °C, 7 d (n = 32). **p**, $tsh^{ts} > lacZ^{RNAi}$, 29 °C, 7 d (n = 22). **q**, $tsh^{ts} > Arf79F^{RNAi}$, 29 °C, 7 d (n = 25). **r**, esg^{ts} $wg^{ts} > Ras^{V12} + \zeta\text{-}Cop^{RNAi}$, 29 °C, 7 d ($n$ = 27). **s**, esg^{ts} $wg^{ts} > N^{DN} + \delta\text{-}Cop^{RNAi}$, 29 °C, 7 d ($n$ = 30). **t**, **u**, $esg^{ts} > N^{DN} + Arf79F^{RNAi}$, 29 °C, 7 d ($n$ = 40). **v**, esg^{ts} $wg^{ts} > Ras^{V12} + Acsl^{RNAi}$, 29 °C, 7 d. (**w**) esg^{ts} $wg^{ts} > \beta\text{-}Cop^{RNAi} + Hnf4$, 29 °C, 7 d. **x**, esg^{ts} $wg^{ts} > \zeta\text{-}Cop^{RNAi} + scu$, 29 °C, 7 d. White arrows in **b**–**j** and **r**, **s**, **w**, **x**, point to the hindgut–midgut junction. Yellow arrows point to $Pros^{+}$ enteroendocrine cells in **t**; green arrows point to Dl^{+} ISCs in **t**, and a white arrow points to a remaining GFP^{+} stem cell in **u**. Scale bars in **b**–**j** and **l**–**x**, 10 μm.



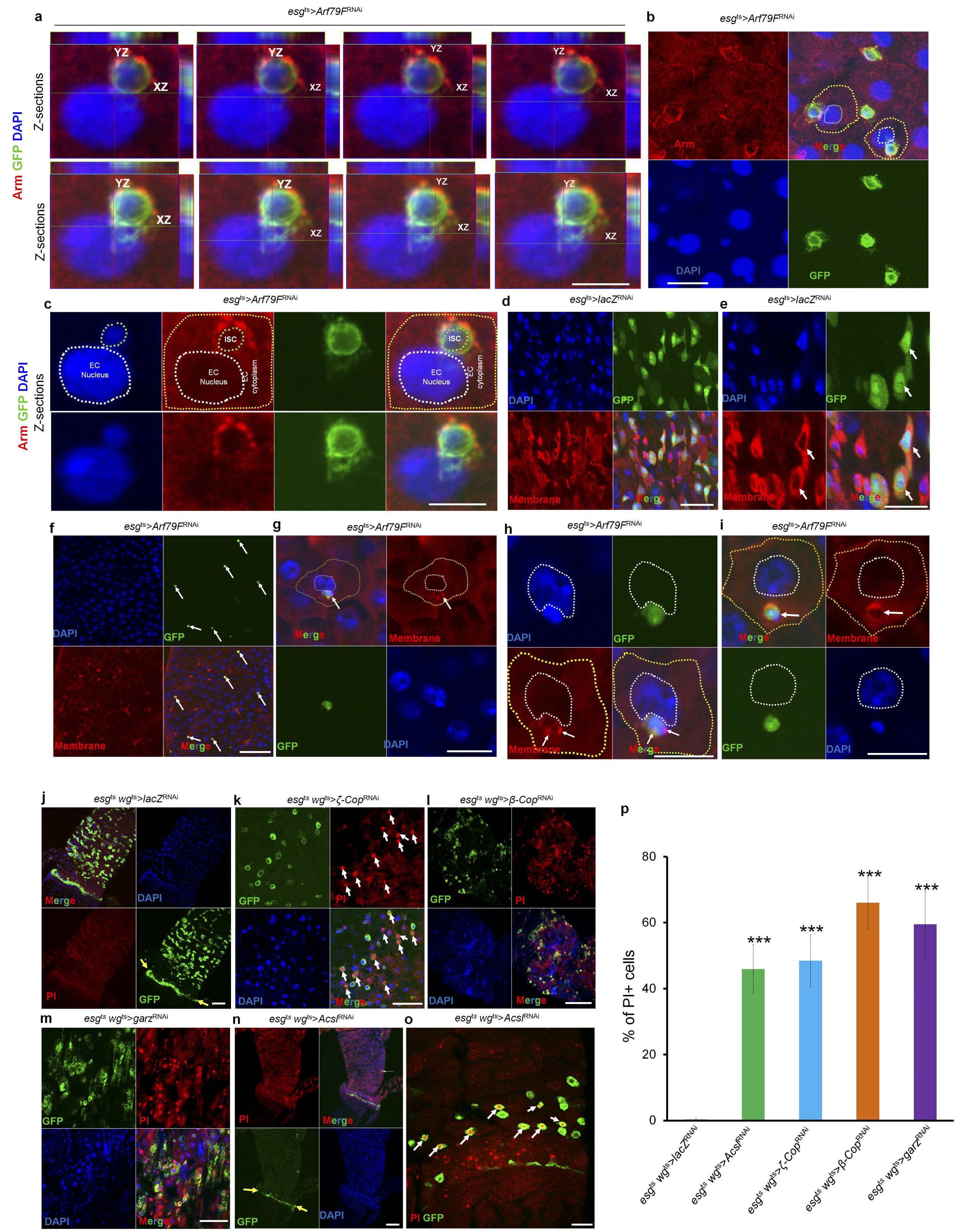
Extended Data Figure 3 | See next page for caption.

Extended Data Figure 3 | The COPI–Arf79F complex regulates stem cell survival through lipolysis and knockdown of these genes blocks lipolysis, but promotes lipid storage. a–h, The COPI complex autonomously regulates stem cell survival. The three- or four-day-old adult female flies were heat-shocked twice with an interval of 8–12 h, at 37 °C, for 60 min to induce MARCM clones[17]. In wild-type clones, small GFP$^+$ cell clusters were detected 2 d ACI (**a, h**; $n = 33$), which grew into large clusters that contained both ISCs and their differentiated progenies by 7 d ACI (**b, h**; $n = 37$). In the *δ-COP* mutant clones, only a few GFP$^+$ cells were identified 2 d ACI (**c, h**; $n = 34$), and none were seen at 7 d ACI (**f**, **h**; $n = 31$). Similarly, only a few GFP$^+$ cells were identified at 2 d ACI in *γ-COP* (**e, h**; $n = 27$) mutant clones, and none were seen at 7 d ACI (**f; h**; $n = 34$). Expressing *UAS–γ-COP–GFP* in *γ-COP*^{10}-mutant MARCM clones (**g** and **h**; $n = 31$) completely rescued the stem cell death phenotype. These results suggest that the COPI complex cell-autonomously regulates stem cell survival. Dotted lines in **a** and **b** outline GFP$^+$ clones. White arrows in **c** and **e** point to individual GFP$^+$ cells. **h**, Quantification of GFP$^+$ cells in the indicated panels. Data show the mean ± s.e.m. Statistical significance was determined by Student's *t*-test, $***P < 0.0001$. The posterior midguts of flies with the indicated genotypes were dissected, stained with the indicated antibodies and analysed by confocal microscopy. **i–p**, The lipolysis pathway is active in stem cells. To further investigate the function of lipolysis in stem cells, we investigated the expression of a lipolysis reporter (*GAL4–dHFN4; UAS–nlacZ*)[21]. In our system, this reporter showed strong β-galactosidase expression in mira-GFP-positive ISCs and RNSCs (**i–k**, $n = 15$), but not in enterocytes, enteroendocrine cells, and the quiescent HISCs of mature adult flies (**i**, white arrows, 3–5 days old) or in the quiescent ISCs of freshly emerged young adult flies (less than 3 days old; **l** and **m**, $n = 17$)[49] at 29 °C culture conditions. Expressing *δ-COP*$^{\mathrm{RNAi}}$ (*esg*$^{\mathrm{ts}}$ > *δ-COP*$^{\mathrm{RNAi}}$ + *GAL4–dHFN4; UAS–nlacZ*) almost completely eliminated the reporter expression (**n**, $n = 24$), suggesting that the reporter is specifically regulated by the COPI complex. We also expressed a constitutively active form of JAK (*hop*$^{\mathrm{Tum\text{-}l}}$) with *GAL4–dHFN4; UAS–nlacZ* and found that the reporter was expressed in *hop*$^{\mathrm{Tum\text{-}l}}$-activated HISCs (**o**, white arrows, $n = 20$). The GAL4 in the reporter system is under the control of an *hsp70* promoter; we heat-shocked the flies for 30 min at 37 °C 12 h before dissection and found that the reporter was strongly expressed in ISCs, RNSCs and HISCs (particularly strong in HISCs), but not in enteroendocrine cells and enterocytes (**p**, white arrows, $n = 17$). Arrows in **i**, **n**, **o** and **p** point to HISCs at the hindgut–midgut junction. **q–v**, Arf79F knockdown promotes lipid storage in stem cells. The genotypes of the flies in each panel were: **q–s**, *esg*$^{\mathrm{ts}}$ > *lacZ*$^{\mathrm{RNAi}}$, 29 °C, 4 d ($n = 30$). **t–v**, *esg*$^{\mathrm{ts}}$ > *Arf79F*$^{\mathrm{RNAi}}$, 29 °C, 4 d ($n = 37$). The posterior midguts of flies with the indicated genotypes were dissected, stained with Oil Red O (red), anti-GFP (green) and DAPI (blue), and analysed by confocal microscopy. Dotted lines outline stem cells and white arrows point to lipid droplets in stem cells. Scale bars in **a–g** and **i–v**, 10 μm.

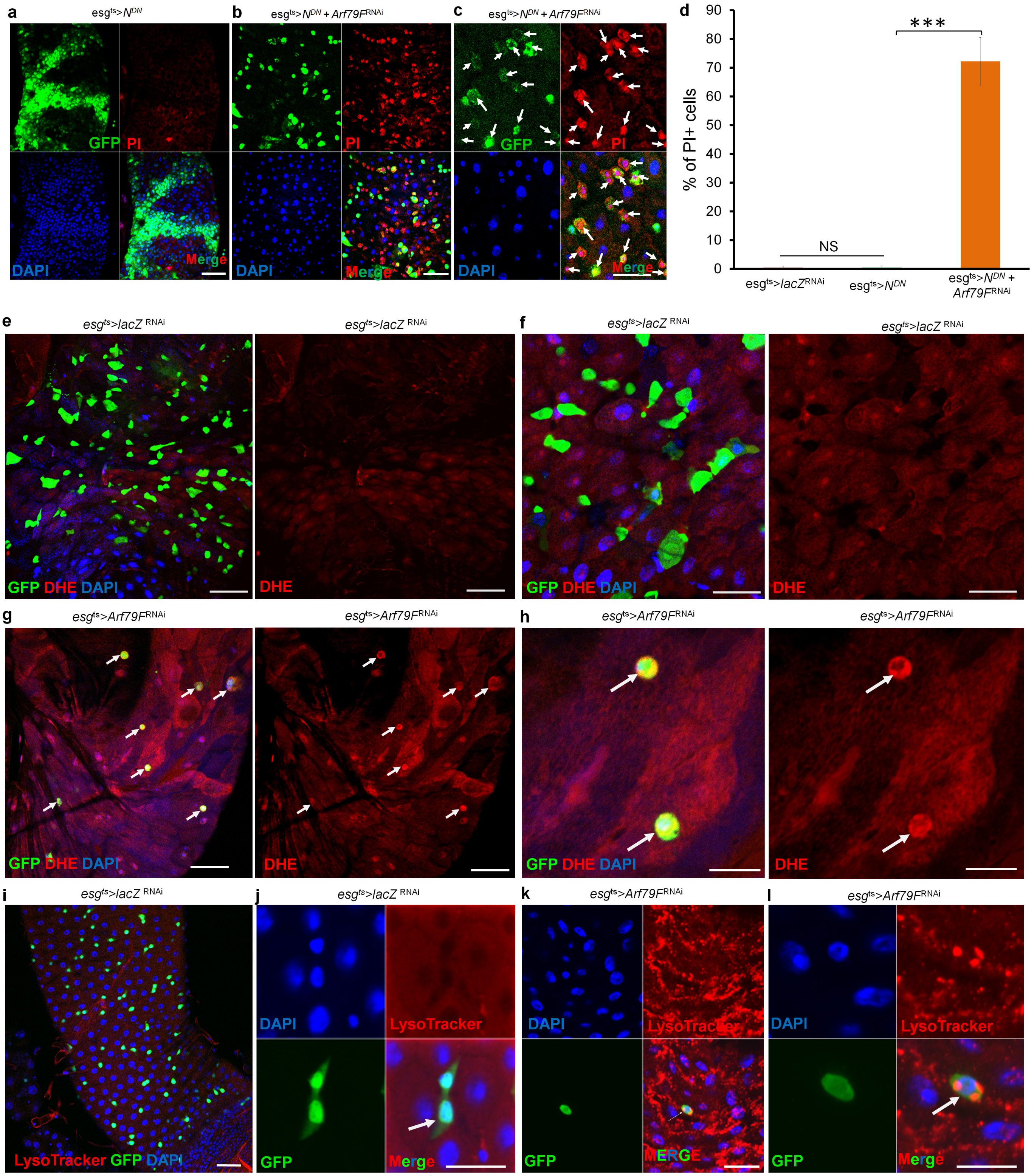


Extended Data Figure 4 | See next page for caption.

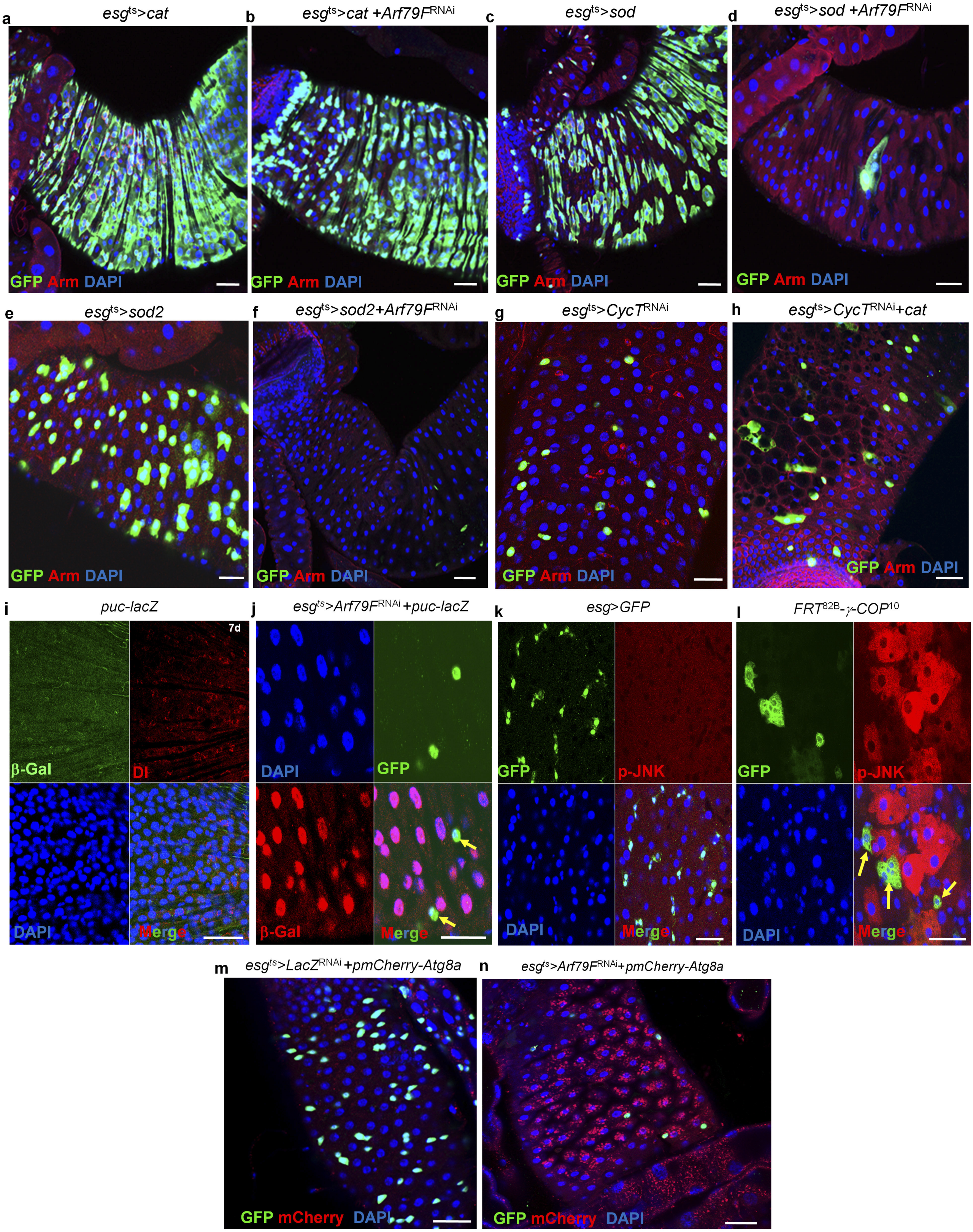
Extended Data Figure 4 | The lipolysis–β-oxidation pathway regulates survival of transformed stem cells. a–**l**, Arf1 inhibitors kill Ras^{V12}-transformed RNSCs through the ROS–Rac–JNK pathway. The GFP-labelled RNSC tumour clusters were induced by expressing Ras^{V12} in RNSC clones, using the positively marked mosaic lineage (PMML) labelling technique[10,50] in adult *Drosophila*. The flies with *Ras*^{V12}-PMML clones were cultured for 4 d at room temperature on normal food to let the tumour grow and then switched to food with indicated drugs for another 4 d. Flies with *Ras*^{V12}-tumours were given normal food with DMSO (**a**), 50 ng ml^{-1} BFA (**b**), 5 μM GCA (**c**), 50 μM LM11 (**d**), 100 μM LG8 (**e**), 50 μM secin H3 (**f**), 50 μM secin H3 + 50 μM JNK inhibitor Sp600125 (**g**), 50 μM secin H3 + 100 μM Rac1 inhibitor (**h**) or 50 μM secin H3 + 10 mM NAC (**i**). **j**,**k**, *esg*^{ts} flies were fed with normal food with either DMSO (**j**, $n = 20$) or 50 μM secin H3 (**k**, $n = 22$). n = number of tissues observed. **l**, Quantification analysis of tumour sizes in Malpighian tubules of indicated panels. We classified all tumours into four categories based on the total number of GFP$^+$ cells in each tumour clone (<10 cells, 10–20 cells, 20–50 cells and 50–100 cells). Total number of tumours investigated for each treatment: DMSO (466 tumours, $n = 27$ Malpighian tubules), BFA (63 tumours, $n = 30$ Malpighian tubules), GCA (73 tumours, $n = 32$ Malpighian tubules), LM11 (94 tumours, $n = 35$ Malpighian tubules), LG8 (86 tumours, $n = 27$ Malpighian tubules), secin H3 (64 tumours, $n = 25$ Malpighian tubules), secin H3 + JNK inhibitor (220 tumours, $n = 30$ Malpighian tubules), Secin H3 + Rac1 inhibitor (211 tumours, $n = 27$ Malpighian tubules), and Secin H3 + NAC (297 tumours, $n = 35$ Malpighian tubules). Arrows point to GFP$^+$ RNSC tumour clusters in **a**–**i**. **m**–**r**, The lipolysis pathway regulates survival of transformed stem cells. The genotypes of the flies in each panel were: **m**, **m′**, *esg*^{ts} > N^{DN}, 29 °C, 4 d (**m**, $n = 30$; **m′**, $n = 35$). **n**, **n′**, *esg*^{ts} > N^{DN} + *Hnf4*, 29 °C, 4 d (**n**, $n = 25$; **n′**, $n = 27$). **o**, **o′**, *esg*^{ts} > *Ras*^{V12}, 29 °C, 4 d (**o**, $n = 25$; **o′**, $n = 32$). **p**, **p′**, *esg*^{ts} > *Ras*^{V12} + *Hnf4*, 29 °C, 4 d (**p**, $n = 25$; **p′**, $n = 32$). **q**, **q′**, *esg*^{ts} > *Ras*^{V12} + *scu*, 29 °C, 4 d (**q**, $n = 23$; **q′**, $n = 30$). The flies were fed with normal food with either DMSO (**m**–**q**) or BFA (**m′** and **n′**, 200 ng ml^{-1}; **o′**–**q′**, 50 ng ml^{-1}) for 4 d. Expressing *Hnf4* or *scu* partially blocked the effect of BFA on transformed stem cells. **r**, Quantification of esg > GFP$^+$ tumour cells in 5×10^3 μm^2 per treatment in indicated panels. Data show the mean ± s.e.m. Statistical significance was determined by Student's *t*-test, ***$P < 0.0001$. NS, not significant ($P > 0.05$). Arrows point to GFP$^+$ RNSC tumour clusters in **o**–**q′**. **s**–**y**, FAO inhibitors, but not 2-DG, kill Ras^{V12}-transformed RNSCs. The GFP-labelled RNSC tumour clusters were induced by expressing Ras^{V12} in RNSC clones using the PMML technique in adult *Drosophila*. The flies with *Ras*^{V12}-PMML clones were cultured for 4 d at room temperature on normal food to let the tumour grow and then switched to food with indicated drugs for another 4 d. Flies with *Ras*^{V12}-tumours were given normal food with DMSO (**s**), 5 μM triacsin C (**t**), 100 μM mildronate (**u**, $n = 27$), 100 μM etomoxir (**v**), 100 μM enoximone (**w**, $n = 37$) or 50 mM 2-deoxyglucose (2-DG) (**x**, $n = 32$). **y**, Quantification analysis of tumour sizes in Malpighian tubules of indicated panels. Total number of tumours investigated for each treatment: DMSO (474 tumours, $n = 30$ Malpighian tubules), triacsin C (47 tumours, $n = 32$ Malpighian tubules), mildronate (69 tumours, $n = 27$ Malpighian tubules), etomoxir (73 tumours, $n = 35$ Malpighian tubules), enoximone (86 tumours, $n = 27$ Malpighian tubules) and 2-DG (264 tumours, $n = 32$ Malpighian tubules). Arrows point to GFP$^+$ RNSC tumour clusters. The gut of flies with the indicated genotypes was dissected after cultured, stained with the indicated antibodies and analysed by confocal microscopy. Scale bars in **a**–**k**, **m**–**q'** and **s**–**x**, 10 μm.



Extended Data Figure 5 | Knockdown of components of the COPI–Arf79F–Acsl pathway kill normal and transformed stem cells through necrosis. **a**–**i**, The genotypes of the flies in each panel were: **a**–**c** and **f**–**i**, $esg^{ts} > Arf79F^{RNAi}$, 29 °C, 7 d ($n = 27$). **d**, **e**, $esg^{ts} > lacZ^{RNAi}$, 29 °C, 7 d ($n = 20$). In **d**–**i**, a dye (CellMask) marks plasma membranes. In **a**-**c** and **g**-**i**, a dying ISC is engulfed by a neighbouring enterocytes. **j**, $esg^{ts}\ wg^{ts} > lacZ^{RNAi}$, 29 °C, 4 d ($n = 30$). **k**, $esg^{ts}\ wg^{ts} > \zeta\text{-}cop^{RNAi}$, 29 °C, 4 d ($n = 36$). **l**, $esg^{ts}\ wg^{ts} > \beta\text{-}Cop^{RNAi}$, 29 °C, 4 d ($n = 34$). **m**, $esg^{ts}\ wg^{ts} > garz^{RNAi}$, 29 °C, 4 d ($n = 32$). **n**, **o**, $esg^{ts}\ wg^{ts} > Acsl^{RNAi}$, 29 °C, 4 d ($n = 32$). **p**, Quantification of propidium-iodide-positive cells in the indicated panels. Data show the mean ± s.d. Statistical significance was determined by Student's *t*-test, $***P < 0.0001$ (compared to control). Yellow arrows point to hindgut–midgut junctions in **j** and **n**, white arrows point to GFP- and propidium-iodide-positive stem cells in **k** and **o**. The posterior midguts of flies with the indicated genotypes were dissected, stained with the indicated antibodies or reagents and analysed by confocal microscopy. Scale bars in **a**–**o**: 10 μm.

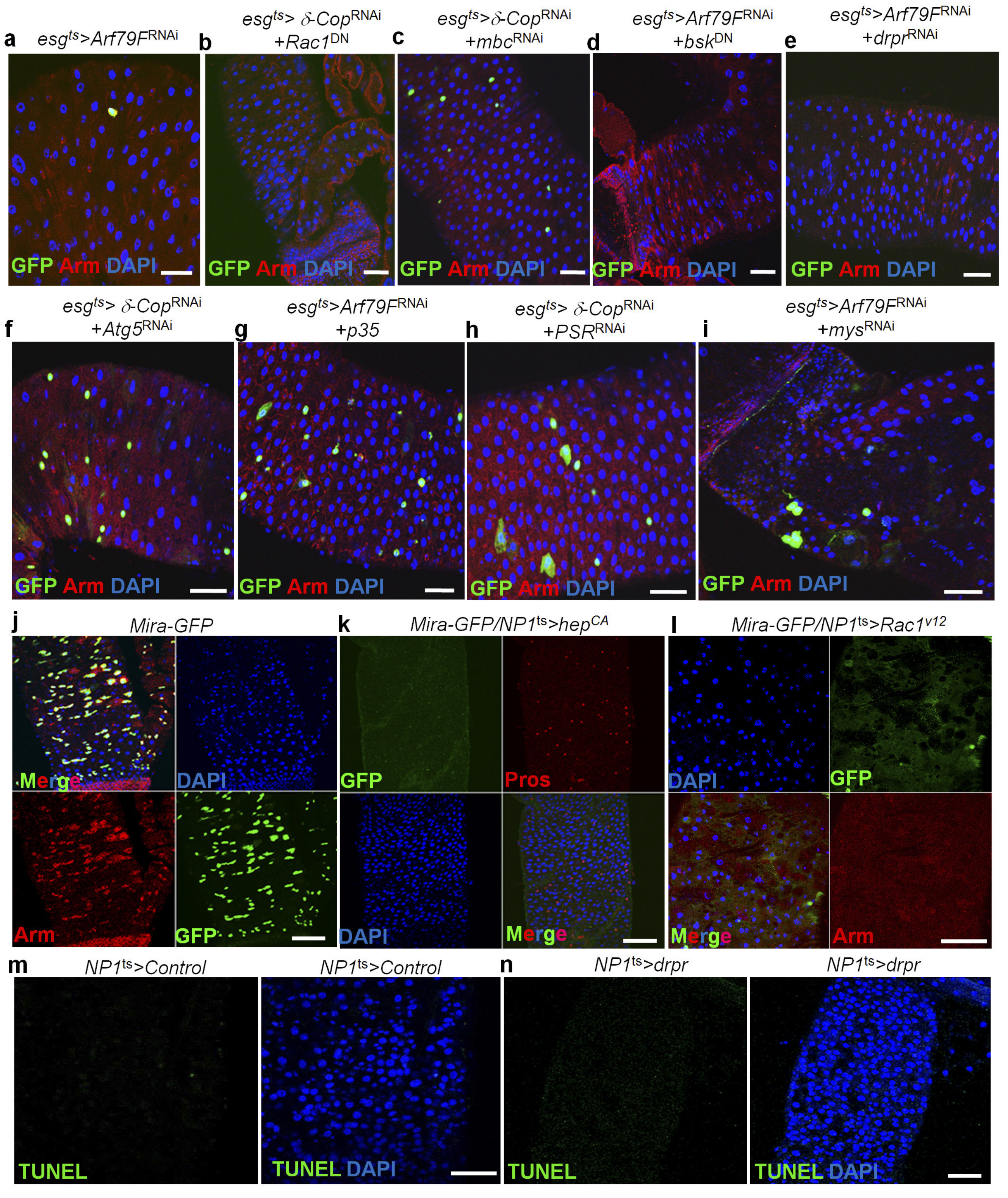


Extended Data Figure 6 | *Arf79F* knockdown kills transformed and normal stem cells through necrosis. (**a**–**d**) The genotypes of the flies in each panel were: **a**, $esg^{ts} > N^{DN}$, 29 °C, 4 d ($n = 25$). **b**, **c**, $esg^{ts} > N^{DN} + Arf79F^{RNAi}$, 29°C, 4 d ($n = 25$). **d**, Quantification of PI^{+} cells in the indicated panels. Data show the mean ± s.d. Statistical significance was determined by Student's *t*-test, $***P < 0.0001$. NS, not significant ($P > 0.05$). White arrows point to GFP- and propidium-iodide-positive stem cells in **c**. **e**–**l**, The genotypes of the flies in each panel were: **e**, **f**, **i**, **j**, $esg^{ts} > lacZ^{RNAi}$, 29 °C, 7 d (**e**, **f**, $n = 27$; **i**, **j**, $n = 30$). **g**, **h**, **k**, **l**, $esg^{ts} > Arf79F^{RNAi}$, 29 °C, 7 d (**g**, **h**, $n = 45$; **k**, **l**, $n = 30$). White arrows point to GFP^{+} stem cells. No DHE or LysoTracker signals were detected in the wild-type midgut, but these signals were intense in the $esg^{ts} > Arf79F^{RNAi}$ flies, indicating that the dying ISCs had accumulated ROS and were intracellularly acidified. The posterior midguts of flies with the indicated genotypes were dissected, stained with the indicated antibodies or dyes and analysed by confocal microscopy. Scale bars in **a**–**c** and **e**–**l**: 10 μm.



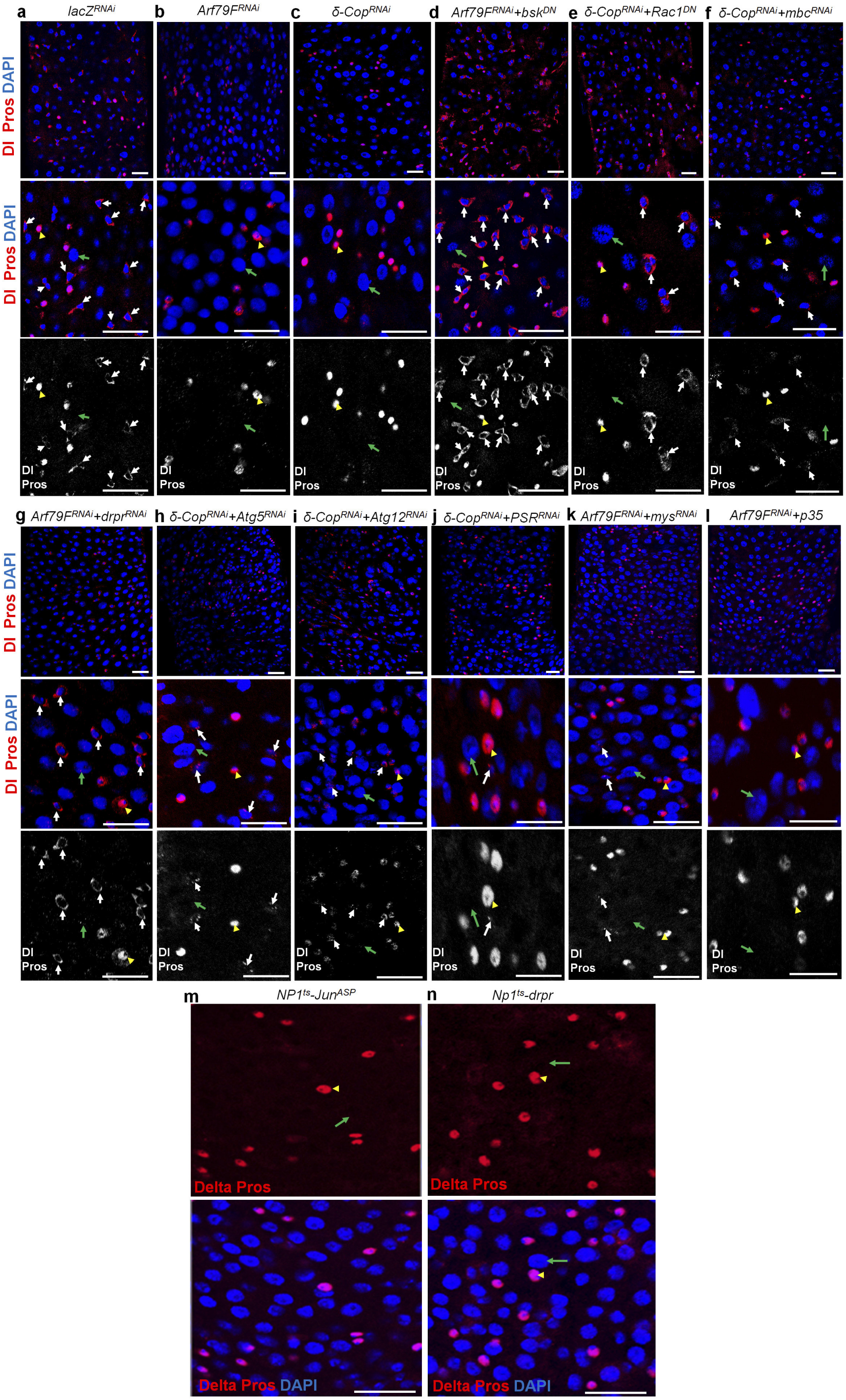
Extended Data Figure 7 | See next page for caption.

Extended Data Figure 7 | Overexpressing *Cat* rescues the ISC death induced by *Arf79F*$^{\mathrm{RNAi}}$ but not *CycT*$^{\mathrm{RNAi}}$ expression and dying ISCs activate JNK signalling and autophagy in enterocytes. **a–h**, The genotypes of the flies in each panel were: **a**, $esg^{ts} > Cat$, 29 °C, 7 d ($n = 25$). **b**, $esg^{ts} > Arf79F^{\mathrm{RNAi}} + Cat$, 29 °C, 7 d ($n = 32$). **c**, $esg^{ts} > Sod$, 29 °C, 7 d ($n = 24$). **d**, $esg^{ts} > Arf79F^{\mathrm{RNAi}} + Sod$, 29 °C, 7 d ($n = 30$). **e**, $esg^{ts} > Sod2$, 29 °C, 7 d ($n = 22$). **f**, $esg^{ts} > Arf79F^{\mathrm{RNAi}} + Sod2$, 29 °C, 7 d ($n = 32$). **g**, $esg^{ts} > CycT^{\mathrm{RNAi}}$, 29 °C, 7 d ($n = 35$). **h**, $esg^{ts} > CycT^{\mathrm{RNAi}} + Cat$, 29 °C, 7 d ($n = 37$). Overexpressing *Cat*, but not *sod* or *sod2*, in stem cells ($esg^{ts} > Arf79F^{\mathrm{RNAi}} + Cat$) rescued the stem-cell death induced by *Arf79F* knockdown but not that induced by *CycT* knockdown ($esg^{ts}>CycT^{\mathrm{RNAi}}+Cat$). **i–n**, The genotypes of the flies in each panel were: **i**, *puc–lacZ*, 29 °C, 7 d ($n = 17$). **j**, $esg^{ts} > Arf79F^{\mathrm{RNAi}} + puc\text{–}lacZ$, 29 °C, 7 d ($n = 20$). Yellow arrows point to GFP^{+} cells. **k**, $esg > GFP$, 29 °C, 4 d ($n = 12$). **l**, FRT^{82B}–γ-*COP*10 MARCM clones, 4 d ($n = 17$). Yellow arrows point to GFP^{+} clones. **m**, $esg^{ts} > lacZ^{\mathrm{RNAi}} + pmCherry\text{–}Atg8a$, 29 °C, 7 d ($n = 22$). **n**, $esg^{ts} > Arf79F^{\mathrm{RNAi}} + pmCherry\text{–}Atg8a$, 29 °C, 7 d ($n = 25$). *Arf79F* knockdown in ISCs induced Puc–lacZ (compare **i** with **j**) and Cherry–Atg8a (compare **m** with **n**) expression in enterocytes, p-JNK was induced in enterocytes in γ-*COP* mutant MARCM clones, compare **k** with **l**). The posterior midguts of flies with the indicated genotypes were dissected, stained with the indicated antibodies and analysed by confocal microscopy. Scale bars in **a–n**: 10 μm.

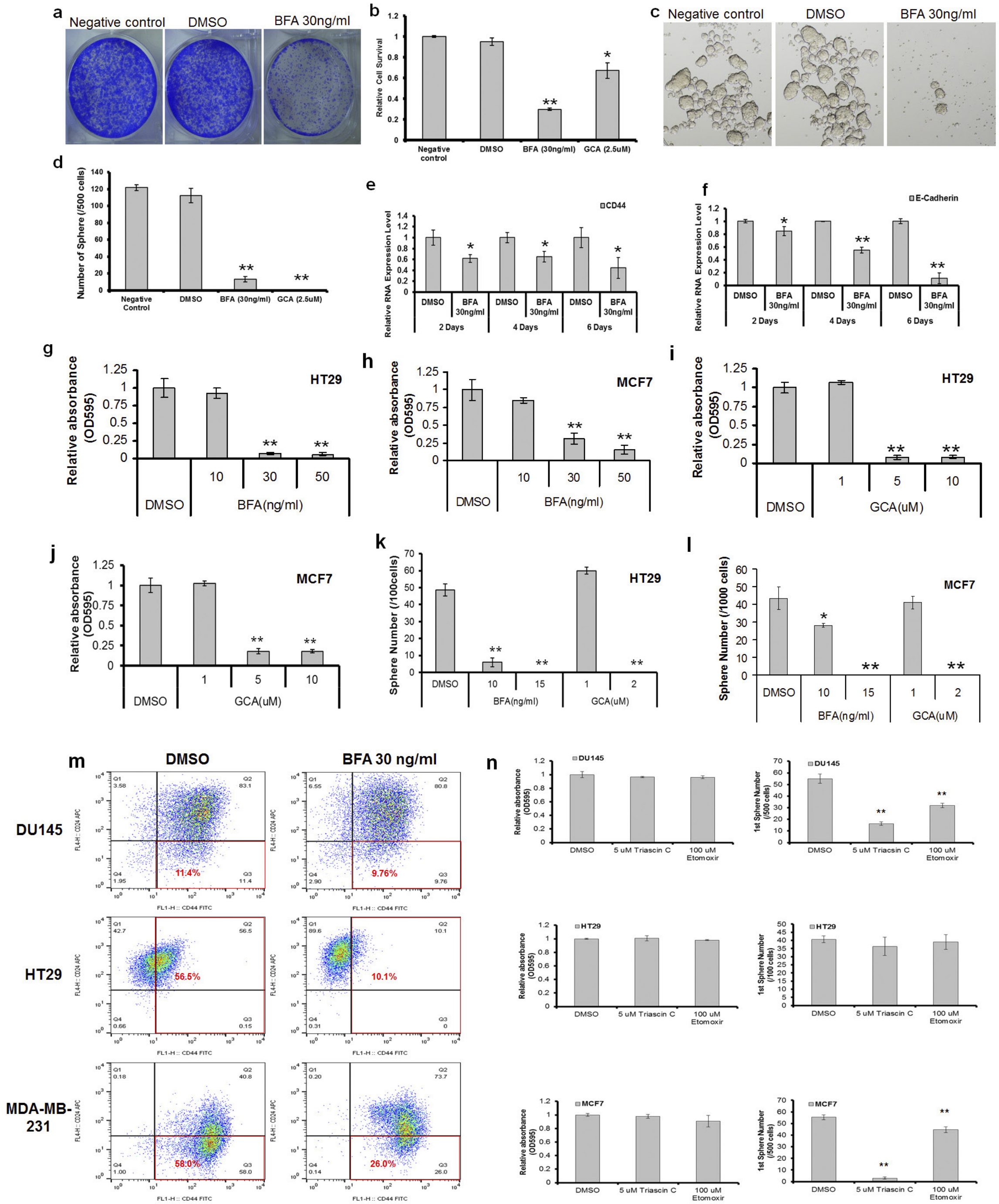


Extended Data Figure 8 | See next page for caption.

Extended Data Figure 8 | Knockdown of components of the JNK pathway or engulfment genes in ISCs did not block the ISC death induced by *Arf79F*$^{\mathbf{RNAi}}$ or *δ-Cop*$^{\mathbf{RNAi}}$ expression. a–i, The genotypes of the flies in each panel were: **a**, *esg*^{ts} > *Arf79F*^{RNAi}, 29 °C, 7 d (*n* = 27). **b**, *esg*^{ts} > *δ-Cop*^{RNAi} + *Rac1*^{DN}, 29 °C, 7 d (*n* = 32). **c**, *esg*^{ts} > *δ-Cop*^{RNAi} + *mbc*^{RNAi}, 29 °C, 7 d (*n* = 25). **d**, *esg*^{ts} > *Arf79F*^{RNAi} + *bsk*^{DN}, 29 °C, 7 d (*n* = 30). **e**, *esg*^{ts} > *Arf79F*^{RNAi} + *drpr*^{RNAi}, 29 °C, 7 d (*n* = 28). **f**, *esg*^{ts} > *δ-Cop*^{RNAi} + *Atg5*^{RNAi}, 29 °C, 7 d (*n* = 32). **g**, *esg*^{ts} > *Arf79F*^{RNAi} + *p35*, 29 °C, 7 d (*n* = 22). **h**, *esg*^{ts} > *δ-Cop*^{RNAi} + *PSR*^{RNAi}, 29 °C, 7 d (*n* = 30). **i**, *esg*^{ts} > *Arf79F*^{RNAi} + *mys*^{RNAi}, 29 °C, 7 d (*n* = 28). *bsk*^{DN} is a dominant-negative form of *Drosophila* JNK (ref. 51), *draper* (*drpr*) encodes a homologue of the *C. elegans* transmembrane phagocytic receptor (ref. 52), *Rac1* encodes a small GTPase that is a homologue of the *C. elegans* engulfment gene *ced-10* (ref. 53), *myoblast city (mbc)/Crk/dCed-12* encodes a Rac1 guanine nucleotide exchange factor (GEF) (ref. 53), *PSR* encodes a phosphatidylserine receptor (ref. 54) and *mys* encodes the β-subunit of integrin, which is involved in mammalian cell engulfment (ref. 55). Light chain 3 (LC3) in autophagosomes is involved in the rapid degradation of the internalized cargo (reviewed in Han and Ravichandran in ref. 27). **j–l**, Activation of *hep* or *Rac1* genes in enterocytes induced the ISC death. The genotypes of the flies in each panel were: **j**, *mira–GFP*, 29 °C, 7 d (*n* = 17). **k**, *mira–GFP* + *NP1*^{ts} (–*UAS–GFP*) > *hep*^{CA} (a constitutively activate form of hep), 29 °C, 7 d (*n* = 15). **l**, *mira–GFP* + *NP1*^{ts} (–*UAS–GFP*) > *Rac1*^{V12} (a constitutively activate form of *Rac1*), 29 °C, 3 d (*n* = 12). **m–n**, Overexpression of *drpr* in enterocytes did not induce EC death. **m**, *NP1*^{ts} > *lacZ*^{RNAi}, 29 °C, 5 d (*n* = 15). **n**, *NP1*^{ts} > *drpr*, 29 °C, 5 d (*n* = 20). The posterior midguts of flies with the indicated genotypes were dissected, stained with the indicated antibodies and analysed by confocal microscopy. Scale bars in **a–n**: 10 μm.



Extended Data Figure 9 | Knockdown of components of the JNK pathway or engulfment genes in enterocytes blocks the ISC death induced by *Arf79F*$^{\mathrm{RNAi}}$ or *δ-Cop*$^{\mathrm{RNAi}}$ expression. The genotypes of the flies in each panel were: **a**, *NP1*$^{\mathrm{ts}}$ *esg*$^{\mathrm{ts}}$ > *lacZ*$^{\mathrm{RNAi}}$, 29 °C, 7 d (n = 20). **b**, *NP1*$^{\mathrm{ts}}$ *esg*$^{\mathrm{ts}}$ > *Arf79F*$^{\mathrm{RNAi}}$, 29 °C, 7 d (n = 32). **c**, *NP1*$^{\mathrm{ts}}$ *esg*$^{\mathrm{ts}}$ > *δ-Cop*$^{\mathrm{RNAi}}$, 29 °C, 7 d (n = 30). **d**, *NP1*$^{\mathrm{ts}}$ *esg*$^{\mathrm{ts}}$ > *Arf79F*$^{\mathrm{RNAi}}$ + *bsk*$^{\mathrm{DN}}$, 29 °C, 7 d (n = 30). **e**, *NP1*$^{\mathrm{ts}}$ *esg*$^{\mathrm{ts}}$ > *δ-Cop*$^{\mathrm{RNAi}}$ + *Rac1*$^{\mathrm{DN}}$, 29 °C, 7 d (n = 202). **f**, *NP1*$^{\mathrm{ts}}$ *esg*$^{\mathrm{ts}}$ > *δ-Cop*$^{\mathrm{RNAi}}$ + *mbc*$^{\mathrm{RNAi}}$, 29 °C, 7 d (n = 18). **g**, *NP1*$^{\mathrm{ts}}$ *esg*$^{\mathrm{ts}}$ > *Arf79F*$^{\mathrm{RNAi}}$ + *drpr*$^{\mathrm{RNAi}}$, 29 °C, 7 d (n = 32). **h**, *NP1*$^{\mathrm{ts}}$ *esg*$^{\mathrm{ts}}$ > *δ-Cop*$^{\mathrm{RNAi}}$ + *Atg5*$^{\mathrm{RNAi}}$, 29 °C, 7 d (n = 17). **i** *NP1*$^{\mathrm{ts}}$ *esg*$^{\mathrm{ts}}$ > *δ-Cop*$^{\mathrm{RNAi}}$ + *Atg12*$^{\mathrm{RNAi}}$, 29 °C, 7 d (n = 25). **j**, *NP1*$^{\mathrm{ts}}$ *esg*$^{\mathrm{ts}}$ > *δ-Cop*$^{\mathrm{RNAi}}$ + *PSR*$^{\mathrm{RNAi}}$, 29 °C, 7 d (n = 22). **k**, *NP1*$^{\mathrm{ts}}$ *esg*$^{\mathrm{ts}}$ > *Arf79F*$^{\mathrm{RNAi}}$ + *mys*$^{\mathrm{RNAi}}$, 29 °C, 7 d (n = 35). **l**, *NP1*$^{\mathrm{ts}}$ *esg*$^{\mathrm{ts}}$ > *Arf79F*$^{\mathrm{RNAi}}$ + *p35*, 29 °C, 7 d (n = 27). **m**, *NP1*$^{\mathrm{ts}}$ > *Jra*$^{\mathrm{Asp}}$ (a constitutively activate form of Jun), 29 °C, 7 d (n = 25). **n**, *NP1*$^{\mathrm{ts}}$ > *drpr*, 29 °C, 7 d (n = 20). Expression of *Jra*$^{\mathrm{Asp}}$ and *drpr* in enterocytes eliminates Dl^+ ISCs. The posterior midguts of flies with the indicated genotypes were dissected, stained with the indicated antibodies and analysed by confocal microscopy. White arrows point to Dl^+ ISCs, yellow arrowheads point to Pros^+ enteroendocrine cells and green arrows point to enterocytes. Scale bars in **a–n**: 10 μm.



Extended Data Figure 10 | Arf1 and FAO inhibitors suppress CSCs in human cancer cell lines. (**a**–**f**) Arf1 inhibitors suppress proliferation and sphere formation in DU145 cells. **a**, **b**, Crystal violet staining was used to detect cell survival after 2 days of treatment with BFA or GCA at the indicated concentrations in DU145 cells. The growth of DU145 cells was strongly inhibited by 30 ng ml^{-1} BFA (**a**, **b**) and 2.5 μM GCA (**b**). We also tested the two inhibitors in tumour sphere formation by cancer cells, a widely used *in vitro* technique for assessing CSC self-renewal capacity[56]. Spheres were cultured with or without BFA or GCA. The two inhibitors also inhibited tumour sphere formation (**c**, **d**). GCA was a weak inhibitor of growth (**b**), but a strong inhibitor of tumour sphere formation (**d**) in DU145 cells, indicating that these inhibitors may specifically target CSCs. The mRNA level of CD44 and E-Cadherin, two potent prostate cancer tumour-initiating cell markers, were reduced by BFA treatment (**e**, **f**). Data show the mean ± s.e.m. Statistical significance was determined by Student's *t*-test, $*P < 0.05$; $**P < 0.01$ (compared to DMSO). **g**–**l**, Arf1 inhibitors suppress proliferation and sphere formation in HT29 and MCF7 cells. Crystal violet staining was used to detect cell survival after 2 days of treatment with BFA or GCA at the indicated concentrations in HT29 and MCF7 cells. The inhibitors reduced the cell survival rate (**g**–**j**). Spheres were cultured with or without BFA or GCA. The inhibitors inhibited sphere formation dramatically (**k**, **l**). Data show the mean ± s.e.m. Statistical significance was determined by Student's *t*-test, $* P < 0.05$; $**P < 0.01$ (compared to DMSO). (**m**, **n**) BFA and FAO inhibitors reduce CSC in DU145, HT29, MCF7 and MDA-MB-231 cells. **m**, Flow cytometry was used to detect cancer stem cell surface makers in DU145, HT29 and MDA-MB-231 cells after 2 days of treatment with BFA. Cell subpopulations enriched with cancer stem cells ($CD44^+$ and $CD24^-$ for DU145 and MDA-MB-231, $CD44^+$ for HT29) were marked with red line. **n**, Crystal violet staining was used to detect cell survival after 2 days of treatment with triascin C or etomoxir using indicated concentrations in DU145, HT29 and MCF7 cells (left). Spheres were cultured with or without triascin C or etomoxir. The inhibitors markedly inhibited sphere formation in DU145 and MCF7 cells (right). Data show the mean ± s.e.m. Statistical significance was determined by Student's *t*-test, $**$, $P < 0.01$ (compared to DMSO).

doi:10.1038/nature19771

XPO1-dependent nuclear export is a druggable vulnerability in *KRAS*-mutant lung cancer

Jimi Kim[1], Elizabeth McMillan[1], Hyun Seok Kim[2], Niranjan Venkateswaran[3], Gurbani Makkar[1], Jaime Rodriguez-Canales[4], Pamela Villalobos[4], Jasper Edgar Neggers[5], Saurabh Mendiratta[1], Shuguang Wei[6], Yosef Landesman[7], William Senapedis[7], Erkan Baloglu[7], Chi-Wan B. Chow[4], Robin E. Frink[8], Boning Gao[8], Michael Roth[6], John D. Minna[8], Dirk Daelemans[5], Ignacio I. Wistuba[4], Bruce A. Posner[6], Pier Paolo Scaglioni[3] & Michael A. White[1]

The common participation of oncogenic KRAS proteins in many of the most lethal human cancers, together with the ease of detecting somatic *KRAS* mutant alleles in patient samples, has spurred persistent and intensive efforts to develop drugs that inhibit KRAS activity[1]. However, advances have been hindered by the pervasive inter- and intra-lineage diversity in the targetable mechanisms that underlie KRAS-driven cancers, limited pharmacological accessibility of many candidate synthetic-lethal interactions and the swift emergence of unanticipated resistance mechanisms to otherwise effective targeted therapies. Here we demonstrate the acute and specific cell-autonomous addiction of *KRAS*-mutant non-small-cell lung cancer cells to receptor-dependent nuclear export. A multi-genomic, data-driven approach, utilizing 106 human non-small-cell lung cancer cell lines, was used to interrogate 4,725 biological processes with 39,760 short interfering RNA pools for those selectively required for the survival of *KRAS*-mutant cells that harbour a broad spectrum of phenotypic variation. Nuclear transport machinery was the sole process-level discriminator of statistical significance. Chemical perturbation of the nuclear export receptor XPO1 (also known as CRM1), with a clinically available drug, revealed a robust synthetic-lethal interaction with native or engineered oncogenic KRAS both *in vitro* and *in vivo*. The primary mechanism underpinning XPO1 inhibitor sensitivity was intolerance to the accumulation of nuclear IκBα (also known as NFKBIA), with consequent inhibition of NFκB transcription factor activity. Intrinsic resistance associated with concurrent *FSTL5* mutations was detected and determined to be a consequence of YAP1 activation via a previously unappreciated FSTL5–Hippo pathway regulatory axis. This occurs in approximately 17% of *KRAS*-mutant lung cancers, and can be overcome with the co-administration of a YAP1–TEAD inhibitor. These findings indicate that clinically available XPO1 inhibitors are a promising therapeutic strategy for a considerable cohort of patients with lung cancer when coupled to genomics-guided patient selection and observation.

Extensive efforts have been directed at the identification of synthetic-lethal targets in *KRAS*-mutant cancers, producing mixed results[2–5]. One obstacle may be the sampling error that is associated with phenotypic diversity among *KRAS*-mutant cancers, both between and within disease lineages. To examine this, we determined the *KRAS* mutation status of 106 non-small-cell lung cancer (NSCLC)-derived cell lines (Supplementary Table 1). We then delineated common deterministic patterns derived from variations seen in whole-genome mRNA expression[6] (Supplementary Table 2). At least eight phenotypic clusters were recovered, with *KRAS*-mutant cell lines present within most of them (Fig. 1a and Extended Data Fig. 1a). In fact, the variation in mRNA expression among *KRAS*-mutant cell lines was equivalent to that of all other cell lines in the panel (Extended Data Fig. 2a). To enrich for detection of bona fide synthetic-lethal genetic interactions with mutant *KRAS*, we therefore selected 12 cell lines, collectively distributed across the range of KRAS-independent background

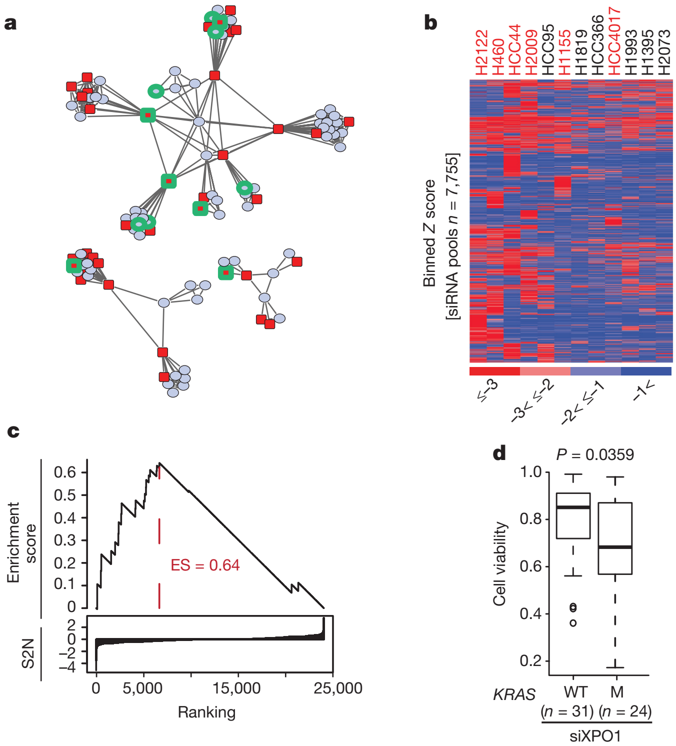


Figure 1 | Synthetic-lethal genetic interactions in *KRAS*-mutant NSCLC cells. a, Two-dimensional APC projection of 106 NSCLC lines based on whole-genome variation in mRNA expression. Nodes represent cell lines and edges represent the Euclidean distance between cell lines. Red nodes, *KRAS* mutant ($n = 37$); blue nodes, *KRAS* wild type ($n = 69$). Cell lines subjected to whole-genome siRNA toxicity screening are highlighted in green. **b**, Binned *Z*-score distributions of selectively toxic gene depletions across the indicated cell lines. Cell lines (in columns; red label, *KRAS* mutant; black label, *KRAS* wild type) and siRNA target genes (rows) are clustered by two-way unsupervised unweighted pair group method with arithmetic mean (UPGMA). **c**, The reactome NEP NS2 interacts with the cellular export machinery. Empirical cumulative siRNA score distribution for a top-ranked *KRAS*-mutant-enriched gene set. S2N, signal-to-noise ratio. **d**, Differences in cell viability following *XPO1* depletion with XPO1 siRNA (siXPO1) in *KRAS*-mutant versus *KRAS*-wild-type (WT) cell lines. Box plots indicate median and interquartile range (IQR). Unpaired *t*-test, $P = 0.0359$.

[1]Department of Cell Biology, UTSW Medical Center, Dallas, Texas 75390, USA. [2]Severance Biomedical Science Institute, Yonsei University College of Medicine, Seoul 120-752, South Korea. [3]Internal Medicine, UTSW Medical Center, Dallas, Texas 75390, USA. [4]Department of Translational Molecular Pathology, MD Anderson Cancer Center, Houston, Texas 77030, USA. [5]KU Leuven Department of Microbiology and Immunology, 3000 Leuven, Belgium. [6]Biochemistry, UTSW Medical Center, Dallas, Texas 75390, USA. [7]Karyopharm Therapeutics, Newton, Massachusetts 02459, USA. [8]Hamon Center, UTSW Medical Center, Dallas, Texas 75390, USA.

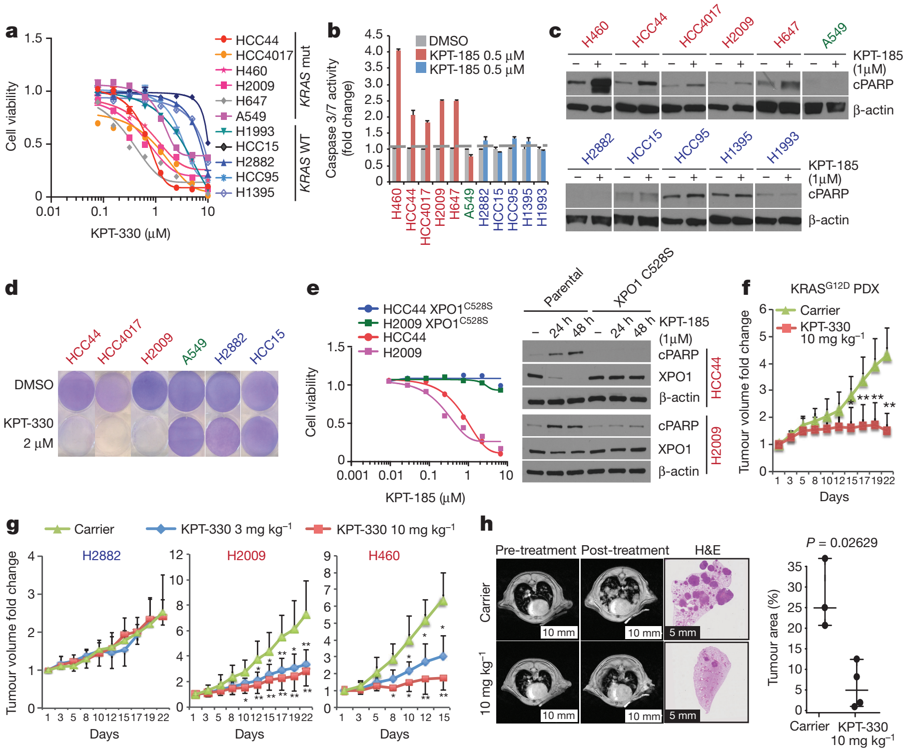


Figure 2 | Selective sensitivity of *KRAS*-mutant NSCLC cells to chemical inhibition of the nuclear transport receptor XPO1. a, Dose–response curves for a panel of NSCLC lines following a 72-h exposure to KPT-330. Data are mean ± s.d. ($n = 3$). **b,** Induction of Caspase 3/7 activity in *KRAS*-mutant lines by KPT-185. Bars indicate mean and range of data ($n = 2$). **c,** Accumulation of the cell death marker cleaved PARP (cPARP) in *KRAS*-mutant lines following exposure to 1 μM KPT-185. β-actin is shown as a loading control. **d,** Stationary-phase cell populations exposed to KPT-330 for 6 d (crystal violet). **e,** Rescue of XPO1 inhibitor toxicity by gene editing. Dose–response curves are as in **a**. Mean and range ($n = 2$). Immunoblots are as in **c**. **f, g,** Fold change in tumour volume in indicated xenografts upon XPO1 inhibition. $*P < 0.05$, $**P < 0.01$, Unpaired *t*-test, mean ± s.d. **h,** Representative lung magnetic resonance image of $Kras^{LSL\text{-}G12D}$;$p53^{fl/fl}$ mouse is shown before and after treatment for each cohort. Post-treatment haematoxylin and eosin (H&E)-stained left lung lobe is shown. Tumour burden was calculated as the tumour area divided by lung area. Unpaired *t*-test, $P = 0.0263$; scale bars, 10 mm (left 4) and 5mm (right 2).

phenotypes, to serve as subjects for whole-genome short interfering RNA (siRNA) toxicity screens[7,8] (Fig. 1a and Extended Data Fig. 1a). We detected 7,755 candidate siRNA pools that reduced the viability of at least one *KRAS*-mutant line (using a *Z* score cut-off of −3; Fig. 1b and Supplementary Table 3). To mitigate noise from 'off-target' siRNA oligonucleotide sequence-specific effects[9], and to account for the complexity of KRAS-independent phenotypic variation, we used gene set enrichment analysis (GSEA; see Methods) to score gene sets, rather than individual genes, with collectively selective activity in the *KRAS*-mutant versus *KRAS*-wild-type lines. Of the 4,725 curated mechanistic gene sets queried, 10 were identified as being significantly enriched within the *KRAS*-mutant cohort (Fig. 1c and Extended Data Fig. 2b; false discovery rate (FDR) < 0.2, $P < 1 \times 10^{-16}$). Leading-edge analysis indicated that multiple genes that encode nuclear transport machinery were common among all 10 gene sets (Extended Data Fig. 2c–e). This enrichment was also seen in retrospective analysis of an independent short hairpin RNA (shRNA) viability screen[3] in an isogenic pair of *KRAS*-wild-type and $KRAS^{G12D}$ colorectal cancer cell lines (Extended Data Fig. 2f). Among the nuclear transport components identified in the siRNA screen, the selective nuclear export receptor XPO1 has been previously identified as druggable[10,11]. We therefore tested the sensitivity to *XPO1* depletion across an additional 55 cell lines and found a strong positive correlation with *KRAS* mutation status (Fig. 1d and Extended Data Fig. 2g, h).

These observations led us to consider selective sensitivity to inhibition of nuclear export as a mutant *KRAS*-associated vulnerability in lung cancer cells. To investigate this, we used the XPO1 inhibitors KPT-185 and KPT-330 (Selinexor)[10,12,13] (Extended Data Fig. 3a). Given the phenotypic enrichment for short doubling times in *KRAS*-mutant cells (Extended Data Fig. 3b and Supplementary Table 4), we selected a test cohort of cells with equivalent proliferation rates. Significant selective dose sensitivity among the *KRAS*-mutant cell lines compared to *KRAS*-wild-type lines was observed with both compounds (Fig. 2a and Extended Data Fig. 3c–e). Sensitivity to XPO1 inhibitors was independent of cell doubling time (Extended Data Fig. 3f and Supplementary Table 4) and was completely rescued by the introduction of the drug-resistant mutation (C528S)[12] using CRISPR/Cas9-induced homologous recombination (Fig. 2e and Extended Data Fig. 3g). The non-responsive A549 cell line was an outlier and was therefore included in all subsequent analyses to represent potential mechanistic exceptions and/or contradictions to the KRAS synthetic-lethal theory.

Sensitivity to XPO1 inhibitors was associated with apoptosis (Fig. 2b, c), which was reversed by the $XPO1^{C528S}$ variant (Fig. 2e). This offered the opportunity to test clearance of stationary-phase cell populations using doses equivalent to those achievable *in vivo* with the orally bioavailable XPO1 inhibitor KPT-330 (ref. 14). With the exception of cell line A549, mutant KRAS-associated bimodal sensitivity to XPO1 inhibitors was evident (Fig. 2d), with preservation of target selectivity at doses over 400% higher than bioactive *in vivo* concentrations (Extended Data Fig. 3h and Supplementary Table 5). Notably, expression of oncogenic KRAS was sufficient to sensitize lung epithelia to XPO1 inhibitors in both proliferative and stationary-phase cultures (Extended Data Fig. 3i). However, cell lines with activating mutations in *NRAS* were not sensitive to XPO1 inhibitors unless they carried a concurrent *KRAS* mutation (Extended Data Fig. 3j). Conservation of efficacy and selectivity *in vivo* was tested and confirmed using three different mouse tumour models: subcutaneous xenograft tumour models with both wild-type and mutant *KRAS* NSCLC lines, a $KRAS^{G12D}$ patient-derived xenograft (PDX) model, and the $Kras^{LSL\text{-}G12D}$,$p53^{fl/fl}$ (p53 is also known as Trp53) genetically engineered mouse (GEM) model (Fig. 2f–h and Extended Data Fig. 3k).

GSEA identified NFκB target genes as being enriched in the XPO1-inhibitor-sensitive cohort (Fig. 3a and Extended Data Fig. 4a), and NFκB target genes were highly overrepresented among the 50 most differentially expressed genes in the sensitive cohort when compared with the resistant cohort (Extended Data Fig. 4b, e). NFκB signalling is often activated by KRAS and can be required for KRAS-driven tumorigenesis[2,15,16]. Notably, XPO1 inhibition resulted in the time-dependent nuclear accumulation of the NFκB negative regulatory protein IκBα (Fig. 3b), inhibition of NFκB promoter activity (Fig. 3d) and inhibition of NFκB target-gene expression (Fig. 3c and Extended Data Fig. 4c, d). The drug-resistant *XPO1* allele cleanly reversed NFκB pathway sensitivity to KPT-185 (Fig. 3d). In addition, *NFKBIA/B* (the gene encoding IκBα/β) depletion was sufficient to confer resistance to XPO1 inhibitors (Fig. 3e and Extended Data Fig. 4f, g).

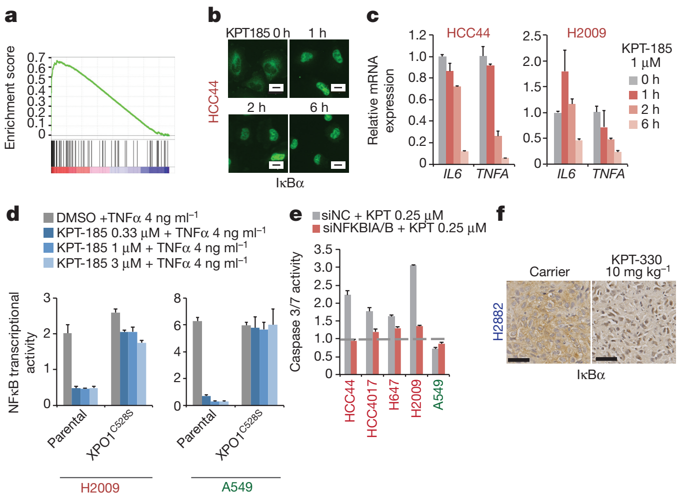


Figure 3 | Selective addiction to NFκB activity specifies sensitivity to XPO1 inhibition. a, GSEA gene set, Hinata NFκB targets keratinocyte up. NFκB transcriptional target enrichment plot (XPO1-inhibitor-sensitive versus XPO1-inhibitor-resistant). **b**, Time-dependent nuclear accumulation of IκBα in response to 1 μM KPT-185. Scale bars, 20 μm. **c**, Time-dependent inhibition of NFκB target gene expression in response to 1 μM KPT-185. Mean and range ($n = 2$). **d**, Rescue of NFκB transcriptional activity by gene editing. Normalized luminescence-based NFκB reporter activity is shown. Mean + s.d. ($n = 3$). **e**, IκB-dependent induction of apoptosis by XPO1 inhibitors. Cells were transfected with the indicated siRNAs and 24 h later they were exposed to 0.25 μM KPT-185 for 48 h. siNC, negative control siRNA. Mean and range ($n = 2$). **f**, Nuclear accumulation of IκBα in xenograft tumours in response to KPT-330 treatment. Scale bars, 50 μm.

These observations suggest that *KRAS*-mutant NSCLC cells require active nuclear export of IκBα to maintain NFκB-dependent survival signalling. Consistent with this, ectopic expression of an IκBα variant with an inactivated nuclear export signal (NES) sequence[17] was tolerated in wild-type but not *KRAS*-mutant NSCLC cells, reducing their viability (Extended Data Fig. 4h). Furthermore, sensitivity to treatment with the weak but specific IκB kinase inhibitor BMS-345541 (ref. 18) exhibited significant positive correlation with sensitivity to KPT-185 (Extended Data Fig. 4i). In contrast, chemical inhibition of MEK activation, a key canonical KRAS pathway effector, had little consequence on cell viability at bioactive concentrations and showed no cooperativity with XPO1 inhibitors (Extended Data Fig. 4j, k). XPO1 inhibitors induced the nuclear accumulation of IκBα in all cell lines tested (Extended Data Fig. 4l), indicating that selective sensitivity is likely to be due to context-specific consequences of inhibition of NFκB signalling rather than selective target inhibition. Consistent with this, we found that KPT-330-resistant tumours displayed extensive nuclear accumulation of IκBα in response to KPT-330 exposure *in vivo* (Fig. 3f). Together, these observations indicate that the oncogenic KRAS protein induces XPO1-dependent activation of NFκB signalling in NSCLC cells to support cell survival, but that activation of the NFκB pathway is not generally required for survival of *KRAS*-wild-type NSCLC tumour lines with alternative mechanistic drivers.

From separate pan-cancer cell line screening efforts, the *KRAS*-mutant NSCLC cell lines H2030 and H2122 were identified and validated as poor responders to XPO1 inhibitors (Extended Data Fig. 5a). We used this finding to identify any potential mechanisms of resistance to XPO1 inhibitors in *KRAS*-mutant NSCLC. On examination of whole-exome sequence data, we found that non-synonymous somatic alterations in *FSTL5* selectively concur in XPO1-inhibitor-resistant *KRAS*-mutant lines (Fig. 4a and Supplementary Table 6). Two previously untested *KRAS*-mutant cell lines, H2291 and H1573, had concurrent *FSTL5* mutations and were both found to be robustly resistant to XPO1 inhibitors (Extended Data Fig. 5b, c). Somatic mutations in *FSTL5* were detected in 10% of lung adenocarcinomas in the The Cancer Genome Atlas (TCGA) database (http://www.cbioportal.org/), with an allelic distribution reminiscent of loss-of-function alterations (Extended Data Fig. 5d). Although mechanistically uncharacterized, FSTL5 has been nominated as a tumour suppressor protein in hepatocellular carcinoma (HCC)[19] (Extended Data Fig. 5e). We found that *FSTL5* depletion selectively reduced sensitivity to XPO1 inhibitors in *KRAS*-mutant, *FSTL5*-wild-type NSCLC lines (Fig. 4b and Extended Data Fig. 5f, g). Furthermore, ectopic FSTL5 expression was tolerated in wild-type but not mutant *FSTL5* cell lines (Extended Data Fig. 5h), suggesting that some cancer genomes place selective pressure on FSTL5 inactivation. This suggests that *FSTL5* mutations detected in cancer cells are loss-of-function and would promote resistance to XPO1 inhibitors.

Defective YAP signalling is a major contributory factor in the development of HCC[20]. Together with observations of YAP-dependent resistance mechanisms to *Kras* inhibition in mouse lung and pancreatic cancers[21–23], this led us to evaluate the potential relationships between FSTL5 and YAP activity. Human lung adenocarcinomas (from the TCGA lung adenocarcinoma database (TCGA-LUAD), $n = 181$) harbouring *FSTL5* somatic alterations displayed significant increases in YAP1 protein expression when compared to wild-type tumours (Fig. 4c). Notably, *FSTL5* depletion was sufficient to induce YAP1 protein stabilization (Fig. 4d and Extended Data Fig. 6a). Furthermore, transcription profiling revealed that the *FSTL5*-dependent gene expression programme was significantly enriched with genes that were also induced upon depletion of the *LATS1* and *LATS2* tumour suppressors (Extended Data Fig. 6b). To evaluate directly the

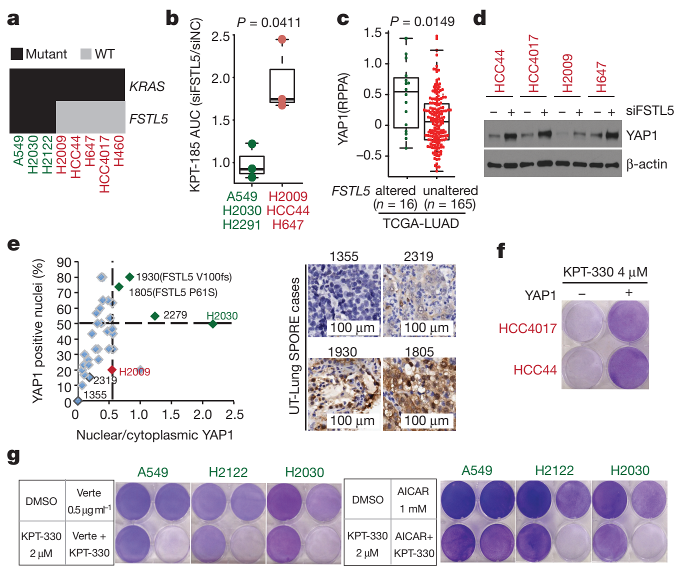


Figure 4 | Concurrent mutations in *FSTL5* are associated with intrinsic resistance of *KRAS*-mutant lines to XPO1 inhibitors and are mechanistically coupled to YAP1 activation. a, Biclustering results for NSCLC cell lines and *KRAS* and *FSTL5* mutation status. **b**, Selective effects of *FSTL5* depletion on the XPO1-inhibitor-sensitivity of *KRAS*-mutant, *FSTL5*-wild-type lines (red) versus *KRAS*-mutant, *FSTL5*-mutant lines (green). Box plot indicates fold changes in area under the curve (AUC) of *FSTL5* siRNA (siFSTL5)-transfected cells normalized to negative control siRNA-transfected cells. AUCs calculated from Extended Data Fig. 5f, g. Unpaired *t*-test, $P = 0.0411$. **c**, Significant enrichment of YAP1 protein in tumours harbouring *FSTL5* somatic alterations (data taken from the TCGA-LUAD). Unpaired *t*-test, $P = 0.0149$. **d**, YAP1 protein accumulation 72 h post-transfection with *FSTL5* siRNAs. **e**, YAP1 immunohistochemistry. Detected somatic *FSTL5* variants are indicated. Representative YAP1 immunohistochemistry stains are shown in the right panel. **f**, Induction of XPO1-inhibitor resistance by YAP1 overexpression (crystal-violet stained). **g**, Induction of XPO1 inhibitors sensitivity by verteporfin/AICAR-mediated YAP pathway inhibition (crystal-violet stained).

FSTL5–YAP relationship in patient samples, 37 *KRAS*-mutant lung adenocarcinoma specimens were immunolabelled with antibodies for YAP1 (Supplementary Table 7). Slides were scored by experienced lung cancer pathologists for the percentage of YAP1-positive tumour cell nuclei and the relative YAP1 nuclear versus cytoplasmic distribution. Comparison to H2009 and H2030 cell blocks revealed three outlier tumours (1805, 1930 and 2279) with predicted YAP pathway activation (Fig. 4e and Extended Data Fig. 6c). Sanger sequencing of *FSTL5* exons identified two samples that harboured somatic non-synonymous *FSTL5* alterations (Fig. 4e and Extended Data Fig. 5d). These observations indicate a strong and clinically relevant association between *FSTL5* mutation status and YAP1 protein accumulation that is consistent with inactivation of the Hippo tumour suppressor pathway[24]. Pertinently, overexpression of YAP1 was sufficient to confer resistance to XPO1 inhibitors (Fig. 4f and Extended Data Fig. 6d, e). Furthermore, chemical (using verteporfin) or genetic (using siRNAs targeted at *YAP*/*TEAD2*) inhibition of YAP1 transcription factor activity was sufficient to confer XPO1 inhibitor sensitivity (Fig. 4g and Extended Data Fig. 6i). Activation of AMPK by the cAMP analogue AICAR can inhibit productive YAP1/TEAD interactions by phosphorylation of YAP1 Ser94 (ref. 25). AICAR also reversed resistance to XPO1 inhibitors, but only in *KRAS*-mutant cell lines with an intact AMPK response (Fig. 4g and Extended Data Fig. 6f–h). Thus, FSTL5 is mechanistically coupled to YAP1 pathway activation, dictating sensitivity of *KRAS*-mutant NSCLC cells to chemical inhibition of XPO1.

Evaluation of the sensitivity of additional NSCLC lines to XPO1 inhibition identified one unexpected *KRAS*-wild-type responder (H1648) and two unexpected *KRAS*-mutant non-responders (HCC515 and Calu1) (Extended Data Fig. 6j). We found that H1648 contains genomic amplification of *IKKB* (a gene encoding IκB kinase β, an inhibitor of IκBα; also known as *IKBKB*) together with a transcription profile indicative of NFκB pathway activation (Extended Data Fig. 6k), suggesting sensitivity to XPO1 inhibitors owing to KRAS-independent addiction to NFκB signalling. HCC515 and Calu1 were found to harbour a *LATS1* mutation (LATS1 R904X) and the loss of Merlin (also known as NF2) expression, respectively (Extended Data Fig. 6l), with both responding to YAP1 inhibition in combination with KPT-330 (Extended Data Fig. 6m). These additional exceptions were therefore accounted for by the mechanistic hypothesis.

Collectively, our observations indicate both that addiction to XPO1-dependent nuclear and cytoplasmic trafficking is a druggable liability in *KRAS*-mutant lung cancers and that genomics-guided patient selection and patient monitoring will be important if maximum benefit is to be achieved from XPO1 inhibitors.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 20 August 2015; accepted 17 August 2016.
Published online 28 September; corrected online 5 October 2016 (see full-text HTML version for details).

1. Cox, A. D., Fesik, S. W., Kimmelman, A. C., Luo, J. & Der, C. J. Drugging the undruggable RAS: mission possible? *Nat. Rev. Drug Discov.* **13,** 828–851 (2014).
2. Barbie, D. A. *et al.* Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature* **462,** 108–112 (2009).
3. Luo, J. *et al.* A genome-wide RNAi screen identifies multiple synthetic lethal interactions with the RAS oncogene. *Cell* **137,** 835–848 (2009).
4. Kumar, M. S. *et al.* The GATA2 transcriptional network is requisite for RAS oncogene-driven non-small cell lung cancer. *Cell* **149,** 642–655 (2012).
5. Scholl, C. *et al.* Synthetic lethal interaction between oncogenic KRAS dependency and STK33 suppression in human cancer cells. *Cell* **137,** 821–834 (2009).
6. Frey, B. J. & Dueck, D. Clustering by passing messages between data points. *Science* **315,** 972–976 (2007).
7. Kim, H. S. *et al.* Systematic identification of molecular subtype-selective vulnerabilities in non-small-cell lung cancer. *Cell* **155,** 552–566 (2013).
8. Whitehurst, A. W. *et al.* Synthetic lethal screen identification of chemosensitizer loci in cancer cells. *Nature* **446,** 815–819 (2007).
9. Jackson, A. L. *et al.* Widespread siRNA "off-target" transcript silencing mediated by seed region sequence complementarity. *RNA* **12,** 1179–1187 (2006).
10. Etchin, J. *et al.* KPT-330 inhibitor of CRM1 (XPO1)-mediated nuclear export has selective anti-leukaemic activity in preclinical models of T-cell acute lymphoblastic leukaemia and acute myeloid leukaemia. *Br. J. Haematol.* **161,** 117–127 (2013).
11. Lapalombella, R. *et al.* Selective inhibitors of nuclear export show that CRM1/XPO1 is a target in chronic lymphocytic leukemia. *Blood* **120,** 4621–4634 (2012).
12. Neggers, J. E. *et al.* Identifying drug-target selectivity of small-molecule CRM1/XPO1 inhibitors by CRISPR/Cas9 genome editing. *Chem. Biol.* **22,** 107–116 (2015).
13. Abdul Razak, A. R. *et al.* First-in-class, first-in-human phase I study of selinexor, a selective inhibitor of nuclear export, in patients with advanced solid tumors. *J. Clin. Oncol.* http://dx.doi.org/10.1200/JCO.2015.65.3949 (2016).
14. Cheng, Y. *et al.* XPO1 (CRM1) inhibition represses STAT3 activation to drive a survivin-dependent oncogenic switch in triple-negative breast cancer. *Mol. Cancer Ther.* **13,** 675–686 (2014).
15. Meylan, E. *et al.* Requirement for NF-κB signalling in a mouse model of lung adenocarcinoma. *Nature* **462,** 104–107 (2009).
16. Bassères, D. S., Ebbs, A., Cogswell, P. C. & Baldwin, A. S. IKK is a therapeutic target in KRAS-induced lung cancer with disrupted p53 activity. *Genes Cancer* **5,** 41–55 (2014).
17. Wuerzberger-Davis, S. M. *et al.* Nuclear export of the NF-κB inhibitor IκBα is required for proper B cell and secondary lymphoid tissue formation. *Immunity* **34,** 188–200 (2011).
18. Burke, J. R. *et al.* BMS-345541 is a highly selective inhibitor of IκB kinase that binds at an allosteric site of the enzyme and blocks NF-κB-dependent transcription in mice. *J. Biol. Chem.* **278,** 1450–1456 (2003).
19. Zender, L. *et al.* An oncogenomics-based *in vivo* RNAi screen identifies tumor suppressors in liver cancer. *Cell* **135,** 852–864 (2008).
20. Fitamant, J. *et al.* YAP inhibition restores hepatocyte differentiation in advanced HCC, leading to tumor regression. *Cell Rep*. **10,** 1692–1707 (2015).
21. Kapoor, A. *et al.* YAP1 activation enables bypass of oncogenic KRAS addiction in pancreatic cancer. *Cell* **158,** 185–197 (2014).
22. Lin, L. *et al.* The Hippo effector YAP promotes resistance to RAF- and MEK-targeted cancer therapies. *Nat. Genet.* **47,** 250–256 (2015).
23. Shao, D. D. *et al.* KRAS and YAP1 converge to regulate EMT and tumor survival. *Cell* **158,** 171–184 (2014).
24. Yu, F.X., Zhao, B., & Guan, K.L. Hippo pathway in organ size control, tissue homeostasis, and cancer. *Cell* **163,** 811–828(2015).
25. Mo, J. S. *et al.* Cellular energy stress induces AMPK-mediated regulation of YAP and the Hippo pathway. *Nat. Cell Biol.* **17,** 500–510 (2015).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank J. Luo for sharing primary data; C. Marcireau for sharing information on merlin expression status in NSCLC lines; C. Xian, S. Zhang, B. Massant, and L.Bral for technical support. This work was supported by grants from the NCI, CPRIT and UT-Lung SPORE.

Author Contributions J.K. and M.A.W. designed the experiments. J.K. performed the experiments with assistance from G.M. E.M. and H.S.K. performed bioinformatic analysis. H.S.K., S.M., S.W., M.R. and B.A.P. conducted the siRNA screen. N.V., P.P.S., B.G. and J.D.M aided preclinical mouse models. R.E.F. and J.D.M. provided the cell doubling times. J.R.-C., P.V., C.-W.B.C. and I.I.W. aided clinical studies. Y.L., W.S. and E.B. provided KPT compounds, advice and bioanalytical data for KPT-330 in mouse plasma. J.E.N. and D.D. generated $XPO1^{C528}$ cells. J.K. and M.A.W. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.A.W. (michael.white@utsouthwestern.edu).

Reviewer Information *Nature* thanks I. Macara, A. Vivancos and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

Cell lines and reagents. NSCLC cell lines were established at the NCI and The University of Texas Southwestern Medical Center or were obtained from the ATCC. They were maintained in RPMI 1640 (Gibco), supplemented with 5% heat-inactivated fetal bovine serum (FBS; Atlanta Biologicals) and 1% penicillin/streptomycin (Gibco) in a humidified chamber at 5% CO_2. HBEC30 and HBEC30-KP were maintained in ACL4 (RPMI 1640 supplemented with 0.02 mg ml^{-1} insulin, 0.01 mg ml^{-1} transferrin, 25 nM sodium selenite, 50 nM hydrocortisone, 10 mM HEPES, 1 ng ml^{-1} EGF, 0.01 mM ethanolamine, 0.01 mM O-phosphorylethanolamine, 0.1 nM triiodothyronine, 2 mg ml^{-1} BSA, 0.5 mM sodium pyruvate) with 2% FBS and 1% penicillin/streptomycin. All cell lines were authenticated using short tandem repeat (STR) profiling (PowerPlex 1.2, Promega) for at least eight different loci and results were compared with reference STR profiles available through the ATCC or established by our laboratory. Following authentication, cell-line stocks were frozen and maintained in liquid nitrogen until they were used in the reported experiments. Polyclonal stable cell lines were established by infecting parental cells with the indicated retroviral vector (Extended Data Figs 5h, 6d) or transfecting parental cells with the indicated plasmid (Extended Data Fig. 2h), followed by antibiotic selection for 7–14 days. $XPO1^{C528S}$ knock-in cell lines were generated by CRISPR/Cas9-induced homologous recombination as previously described[12]. All cell lines were mycoplasma-tested before experiments (iNtRON biotechology).

KPT-185 and KPT-330 were provided by Karyopharm Therapeutics. BMS-345541 and Verteporfin were purchased from Sigma. Antibodies were purchased from Cell Signaling (cPARP 9541, IκBa 4814, YAP1 12395, LATS1 3477, Histone H3 9715, ACC 3676, pACC 3661, Lamin A/C 4777), Sigma (β-actin A1978, Flag F1804), Abcam (Merlin ab88957) and Santa Cruz Biotechnology (XPO1 sc5595 and YAP1 sc101199 for immunofluorescence assay). The mutations in the NES of *NFKBIA* (M45A, L49A, I52A; Extended Data Fig. 4h) and the silent mutations in *XPO1* (<u>T</u>GA<u>C</u>AC<u>G</u>GA<u>C</u>TC<u>A</u>AT<u>T</u>AA<u>C</u>; Extended Data Fig. 2h) were generated using the Q5 Site-Directed Mutagenesis Kit (New England Biolabs). All siRNAs used for small-scale experiments were obtained from Dharmacon. *LONRF1* siRNA was used as a negative control siRNA. All siRNA sequences are provided in Supplementary Table 8.

siRNA screens and data processing. Two commercial genome-wide siRNA libraries from Ambion (library 1, 21,585 genes) and Dharmacon (library 2, 18,175 genes) were purchased in the 96-well plate format. siRNAs were dissolved in siRNA buffer (Dharmacon) overnight to a final concentration of 10 μM and stored at −80 °C before use. Libraries 1 and 2 are a mix of 3 and 4 individual siRNA oligonucleotides per gene, respectively. Transfection protocols were optimized for each cell line as previously described[7]. For reverse transfection, 3 μl of each siRNA pool (10 μM) was transferred to serum-free RPMI (95 μl per well) in empty 96-well assay plates (Costar). 30 μl of this siRNA solution was transferred to an empty 96-well optical assay plate (BioMek), incubated for 5 min, then mixed with 10 μl transfection reagent solution (0.13 μl RNAi Max (Invitrogen) in 10 μl serum-free RPMI), and incubated for 15 min. Cells were collected and diluted in parallel, then added to the siRNA–lipid mix and incubated for 96 h. All screens were performed using biological triplicates. CellTiter-Glo (Promega) assays were performed using 15 μl reagent per well followed by a 10 min incubation before quantitation of luminescence with an Envision plate reader (PerkinElmer). siUBB (siRNA against *UBB*, ubiquitin B; Dharmacon) was used as a positive control for toxicity for all cell lines. Screen data were row- and column-median-normalized and log_2-transformed. Mean values from triplicates were used to calculate batch-centred Z scores using siMacro[26].

Hierarchical clustering by UPGMA was performed using the 'stats' package in R, based on Euclidean distance using the 'complete' agglomeration method.

Functional GSEA analysis. Within each individual cell line, minimum gene-level Z scores were binned according to the following rules:

$$Z \leq -3 \rightarrow 3;\ -3 < Z \leq -2 \rightarrow 2;\ -2 < Z \leq -1 \rightarrow 1;\ Z > -1 \rightarrow 0.$$

GSEA was then performed with the signal-to-noise ranking metric to determine gene sets that contained significantly lower Z scores in *KRAS*-mutant compared to the *KRAS* wild-type cells. A plot of the running sum (Fig. 1c) and the resulting signal-to-noise ratio at each point in the ranked list was constructed in R. The top gene sets preferentially 'lower' (that is, containing genes corresponding to siRNA pools with low (toxic) Z scores) in the *KRAS*-mutant cell lines were defined as those that had a $P < 1 \times 10^{-16}$ and FDR < 0.2 (Extended Data Fig. 2b). We performed a leading-edge analysis using the Broad GSEA software to identify genes enriched across multiple significant gene sets (Extended Data Fig. 2c).

Gene expression and data processing. Raw Illumina HumanWG-6 v3.0 Expression BeadChip files for the NSCLC cell lines used in this study are available from the Gene Expression Omnibus using accession number GSE32036. Data were background-corrected using the 'MBCB' package in R, which provides a model-based background correction method similar to an RMA correction with affymetrix arrays. Data were then quantile-normalized to produce equivalent expression distributions amongst cell lines.

To evaluate the distribution of expression variation within the NSCLC panel, standard deviations were calculated for expression of each of 25,235 genes (Illumina HumanWG-6 v3.0) across the full panel of 106 NSCLC lines, the 37 *KRAS*-mutant NSCLC lines, and the 69 *KRAS*-wild-type NSCLC lines (Supplementary Table 1, 2 and refs 27–30). Kernel density estimates were determined using the 'stats' package in R.

To examine the gene regulatory pathways affected by XPO1 inhibition, cells were exposed to either DMSO or 1 μM KPT-185 for 12 h. Total mRNA was isolated and gene expression profiling was performed using Illumina HT12v4 BeadChip. Expression values were extracted using GenomeStudio 2010.2. The raw values were background-corrected, quantile-normalized, log_2-transformed and subjected to GSEA.

To examine the transcriptional response to *LATS* and *FSTL5* depletion, cells were first transfected with siLATS1/2 and siFSTL5. 72 h post-transfection, cells were processed for gene expression profiling as described above. The raw intensities were background-corrected, quantile-normalized and log_2-transformed. Genes with log_2 expression values ≤4 across the samples were excluded from further analysis.

For targeted gene expression analysis, total cellular RNA was isolated using RNeasy miniprep Kit (Qiagen). cDNA was then synthesized using High-Capacity RNA-to-cDNA kits (Applied Biosystems) and subjected to quantitative PCR (qPCR) with TaqMan gene expression assay kits (Applied Biosystems).

Unpaired t-tests and two-sample Kolmogorov–Smirnov tests were performed using the R 'stats' package.

Affinity-propagation-based similarity clustering analysis. Clustering analysis was performed with the affinity propagation clustering (APC) algorithm using the 'apcluster' package in R. APC is a deterministic clustering method that identifies the number of clusters and cluster 'exemplars' (that is, the cluster centroid or the data point that is the best representative of all the other data points within that cluster) entirely from the data[6], giving it an advantage over non-deterministic methods subject to a biased randomized initialization step, such as hierarichial clustering or methods in which the number of clusters has to be pre-specified, such as k-means clustering.

APC performs clustering by passing messages between the data points. It takes as input a square matrix representing pairwise similarity measures between all data points. The algorithm views each data point as a node in a network and is initialized by connecting all the nodes together, where edges between nodes are proportional to Euclidean distance. The algorithm then iteratively transmits messages along the edges, pruning edges with each iteration until a set of clusters and exemplars emerges.

Two real-valued messages are passed between nodes. The 'responsibility' message computes how well-suited point i is to choose point k as an exemplar, given all the other candidate exemplars, k', and is updated by:

$$r(i,k) \leftarrow s(i,k) - \max_{k' st\ k' \neq k}\{a(i,k') + s(i,k')\}$$

The availability message, $a(i,k)$, computes how appropriate it is for point i to select point k as an exemplar, taking into account all the other points for which k is an exemplar, i', and is given by:

$$a(i,k) \leftarrow \min\left\{0, r(k,k) + \sum_{i' st\ i' \notin \{i,k\}} \max(0, r(i',k)\right\}$$

In the above equation, $a(i, k)$ is set to the self-responsibility, $r(k, k)$, plus the sum of the positive responsibilities candidate k receives from other points. The entire sum is thresholded at 0, with a negative availability indicating that it is inappropriate for point i to choose point k as an exemplar so the tie is severed. The self-availability, $a(k, k)$, reflects the accumulated evidence that point k is an exemplar and is updated with the following rule, which reflects the evidence that k is an exemplar based on the positive responsibilities sent to k from all points, and is updated by:

$$a(k,k) \leftarrow \sum_{i' st\ i' \notin \{i,k\}} \max(0, r(i',k)$$

In the first iteration, all points are considered equally likely to be candidate exemplars, and $a(i, k)$ is set to 0 and $s(i, k)$ is set to the input similarity measure

between points i and k. The above rules are then iteratively updated until a clear, stable set of clusters and exemplars emerges.

In our implementation, we first used the algorithm to identify an initial set of exemplars and clusters from the data matrix. The exemplars were then clustered together and this procedure was repeated until no more clusters emerged, identifying a hierarchical structure of clusters. Networks were drawn with cytoscape[31] in the following manner. All members of the primary clusters are interconnected, and edge lengths are drawn to be proportional to Euclidean distances. Edge lengths between exemplars that cluster together are also drawn to be proportional to Euclidean distances. The entire map was rendered in a two-dimensional display using a cytoscape built-in spring-embedded algorithm.

To cluster 106 NSCLC cell lines with defined *KRAS* status by similar expression profiles (Fig. 1a, Extended Data Fig. 1a), we first reduced the panel of genes to those that were expressed at a $\log_2$-normalized expression value of 6 in at least one cell line and those that were present in the top 20% of the most highly variant genes; this resulted in a panel of 3,101 detectable and variable genes.

Retrospective analysis of genome-wide synthetic-lethal shRNA screen data. Extended Data Fig. 2f was generated using genome-wide synthetic-lethal shRNA screening data[3]. The shRNA screen was performed in an isogenic pair of *KRAS* wild-type and *KRAS*^{G12D} colorectal cancer cell lines with 6 pools of ~13,000 shRNAs/pool targeting 32,293 human transcripts. $\log_2$ fold changes in relative abundance of each shRNA depleted over time were analysed for each sample. Those with a fold-depletion equal or less than 0 in any sample were then compared between samples (*KRAS*-mutant versus *KRAS* wild-type) using the non-parametric two-sample Kolmogorov–Smirnov test.

Cell viability and cytotoxicity assay. For dose–response curves, cells were plated at 50% density in 96-well assay plates. On the following day, serially diluted compounds or vehicle alone were added to the culture media. Cell viability was measured using CellTiter-Glo (Promega) 72 h post-treatment. For dose–response analysis of XPO1 inhibitors combined with either siRNAs or chemical inhibitors (Extended Data Figs 4g, k, 5f, g, 6i), cell viability was normalized to the indicated matching controls. AUCs were computed by the trapezoidal method using GraphPad software. Caspase enzymatic activity was analysed using Caspase-Glo 3/7 (Promega) after compound treatment according to manufacturer's instructions. The raw luminescence values were divided by the average luminescence values of matching controls (Figs 2b, 3e). To examine the cytotoxic effect of compounds on post-confluent cells, NSCLC cells were cultured to confluence in 6-well plates, exposed to compounds as indicated, then fixed in 100% cold methanol for 10 min and stained with 0.5% crystal violet for 30 min at room temperature. To examine tolerance to ectopic expression of nuclear IκBα, test plasmid DNA (pEGFP-C3-IκBα-NES-Mut) and control plasmid DNA (pEGFP-C3) were transfected into H2882 cells and H2009 cells in 12-well plates. 48 h and 72 h after transfection, cells were fixed and the GFP/DAPI ratio was examined using images taken with a Zeiss Plan 20×/0.30 PH1 objective on a Zeiss Axioplan 2E microscope. The following formula was used:

$$\text{Sensitivity to nuclear I}\kappa\text{B}\alpha = \left(\frac{\text{GFP}-\text{IkB}\alpha-\text{NES}-\text{Mut}}{\text{DAPI}}\right) \Big/ \left(\frac{\text{GPF}-\text{empty}}{\text{DAPI}}\right)$$

Proliferation rate measurement of NSCLC lines. Cells were counted at seeding, allowed to grow to a confluence of 80–90% and then harvested and the total cell number was determined. Population-doubling time was calculated using the following formula: number of hours from seeding to collection/$((\log_{n(t)} - \log_{n(t0)})/\log_2)$. n(t) is the number of cells at time of passage and $n(t_0)$ is the number of cells seeded at previous passage[32].

Targeted siRNA and plasmid DNA transfection. For transfection in 96-well plates, 1 µl siRNA (10 µM) in 30 µl of serum-free RPMI was mixed with 0.4 µl of RNAi Max (Invitrogen) in 10 µl of serum-free RPMI. Following a 15 min incubation, the siRNA–lipid mix was transferred to empty 96-well assay plates followed by delivery of single cell suspensions (100 µl per well). For transfection in 6-well plates, 10 µl siRNA (10 µM) in 250 µl of serum-free RPMI was mixed with 7 µl RNAi Max in 250 µl of serum-free RPMI and were delivered to plates followed by delivery of single-cell suspensions (2 ml per well). For plasmid DNA transfection in 12-well plates, 0.5 µg of plasmid DNA in 25 µl of serum-free media was mixed with 1.5 µl Fugene 6 (Promega) in 25 µl of serum-free media. After a 15-min incubation, suspended cells (1 ml per well) were added to the plate with DNA–Fugene 6 complexes. For plasmid DNA transfection in 60-mm dishes, cells were pre-plated and 2 µg DNA/6 µl Fugene 6 complexes in 100 µl of serum-free media were delivered to the cells the next day.

NFκB transcriptional activity reporter assay. Cells were reverse-transfected in 96-well microtitre plates with a reporter plasmid (pGL4.32[luc2p/NFκB-RE/Hygro], Promega) expressing firefly luciferase under the control of a multimerized NFκB-responsive element together with the pRL-SV40 Renilla luciferase control reporter plasmid at a ratio of 9:1. 24 h post-transfection, cells were exposed to compounds as indicated for 24 h and then treated with 4 ng ml^{-1} TNF for 6 h. Luciferase activities were measured using the Dual luciferase reporter system (Promega) according to manufacturer's instructions.

Mouse xenografts. NOD/SCID female mice at 4–9 weeks of age were injected subcutaneously with H2882 (3×10^5), H2009 (5×10^5), and H460 (2.5×10^5) cells. When tumours reached 100 mm^3 or larger, mice were randomly assigned to three cohorts and orally treated with vehicle (0.6% (w/v) PVP K-29/32 and 0.6% (w/v) Pluronic F-68), 3 mg kg^{-1} KPT-330, or 10 mg kg^{-1} KPT-330 three times a week by oral gavage. Tumour volume was monitored with digital callipers using the following formula: Width^2 × Length/2. Nutri-Cal (Tomlyn) was provided to mice throughout the experiment as a nutritional supplement. Mice were killed when xenografts reached 2 cm^3. The number of mice per cohort was as follows: for H2882 xenografts, vehicle, $n=7$; 3 mg kg^{-1}, $n=7$; 10 mg kg^{-1}, $n=6$; for H2009 xenografts, vehicle, $n=8$; 3 mg kg^{-1}, $n=8$; 10 mg kg^{-1}, $n=9$; for H460 xenografts, vehicle, $n=5$, 3 mg kg^{-1}, $n=4$, 10 mg kg^{-1}, $n=5$.

For bioanalysis of KPT-330 in mouse plasma, plasma samples were collected by cardiac puncture from mice ($n=4$ per group), processed with three volumes of methanol containing internal standard (propranolol) and then centrifuged at 1,000*g* for 10 min. The supernatant was analysed by LC–MS/MS using an Agilent 6410 mass spectrometer coupled with an Agilent 1200 HPLC and a CTC PAL chilled autosampler, all controlled by MassHunter software (Agilent). After separation on a C18 reverse-phase high-performance liquid chromatography column (Agilent), using an acetonitrile-water gradient system, peaks were analysed by mass spectrometry using ESI ionization in MRM mode. All plasma samples were compared to a calibration curve prepared in mouse blank plasma.

***Kras*$^{LSL\text{-}G12D}$;*p53*$^{fl/fl}$ GEM model.** *Kras*$^{LSL\text{-}G12D}$;*p53*$^{fl/fl}$ mice were infected with 2.5×10^7 PFU of Ad-Cre (University of Iowa, Gene Transfer Vector Core) intratracheally at 7–8 weeks of age as previously described[33]. Eight to ten weeks after the injection, tumour development was monitored by magnetic resonance imaging (MRI) before treatment. Mice with equivalent tumour burden were randomly assigned to two cohorts and orally treated with either vehicle (0.6% (w/v) PVP K-29/32 and 0.6% (w/v) Pluronic F-68; $n=3$) or 10 mg kg^{-1} KPT-330 ($n=4$) three times a week by oral gavage. Outlier animals presenting with exceptionally high tumour burden were treated with 10 mg kg^{-1} KPT-330 five times a week by oral gavage. After the three weeks of treatment, lungs were imaged by MRI, then collected, fixed with 10% formalin, paraffin-embedded, and H&E stained. The stained lung-tissue specimens were scanned in Hamamatsu Nanozoomer 2.0HT for visualization and evaluation. Tumour burden was determined on the largest lobe and calculated as the tumour area divided by lung area. Area of interest was quantified using ImageJ software.

All magnetic resonance images were obtained using a 7T small-animal MRI scanner (Agilent (Varian), Inc.) equipped with a 40 mm Millipede RF coil (ExtendMR LLC, Milpitas, CA). All MRI acquisitions were gated using both cardiac and respiratory triggering. The images were recorded on the transverse plane, with the major parameters as follows: repetition time (TR), 200; echo time (TE), 1.834 ms; flip angle (FA), 45°; number of average, 8; field of view (FOV), 32 × 32 mm^2; matrix size, 256 × 256; slice number, 17; slice thickness, 1 mm without any gap.

Patient-derived xenografts. Human *KRAS*^{G12D} T2aN0Mx stage lung adenocarcinoma tissue was obtained from a 40-year-old patient with lung cancer and directly implanted into the liver capsule of a NOD/SCID mouse. Written informed consent was obtained from the patient. Tissue procurement for the generation of patient-derived xenografts was approved by the Institutional Review Boards of UTSW. The mouse was killed 18 weeks later and the engrafted tumour was collected, divided, frozen in 10% DMSO/90% FBS and stored at −80 °C until being re-implanted subcutaneously to another NOD/SCID mouse in both flanks. When the tumours reached ~10 mm in length they were resected, evenly divided into 19 pieces and re-implanted subcutaneously to 10-week-old NOD/SCID female mice. Mice harbouring palpable engrafted tumours (58–101 mm^3) were randomly assigned to receive either carrier (0.6% (w/v) PVP K-29/32 and 0.6% (w/v) Pluronic F-68; $n=6$) or 10 mg kg^{-1} KPT-330 ($n=6$) three times a week. Tumour volume was monitored with digital callipers using the following formula: tumour volume = width^2 × length/2.

All mouse studies were performed according to the guidelines of the UT Southwestern Institutional Animal Care and Use Committee.

Immunofluorescence and immunohistochemistry. For immunofluorescence-based imaging, cells were fixed with 3.7% formaldehyde (Fisher Scientific), permeabilized with 0.1% Triton X-100, blocked with PBST (PBS containing 1% bovine serum albumin and 0.1% Tween-20) and incubated with antibodies in PBST against the indicated proteins. Representative images were captured with a PCO: sCMOS 5.5 camera on a Zeiss Axioplan 2E microscope.

For immunohistochemistry-based protein expression analysis, paraffin-embedded mouse tumour samples were deparaffinized, subjected to heat-induced

antigen retrieval in 10 mM sodium citrate buffer, blocked using 3% peroxidase (Sigma), Avidin/Biotin blocking Kit (Vector laboratories), and M.O.M. Kit (Vector laboratories) and incubated with anti-IκBα antibody (Cell Signaling, 1:50) and then biotinylated secondary antibody followed by ABC reagent (Vector Laboratories). The samples were stained using immPACT DAB (Vector Laboratories) and counter stained using Mayer's haematoxylin solution (Sigma).

In tissue microarrays with patient-derived lung tumour samples, immunohistochemistry reactions were performed using a Leica Bond Max automated stainer (Leica Biosystems, Nussloch GmbH). In summary, the NSCLC TMA slides were deparaffinized and hydrated in the Leica Bond autostainer. The primary antibodies used were YAP1 (rabbit monoclonal, clone D8H1X, Cell Signaling Technology, 14074, 1:100). Antigen retrieval was performed using Bond Epitope Retrieval Solution No. 1 (AR9961, Leica Biosystems; equivalent to citrate buffer pH 6.0). The immunohistochemistry reaction was detected using the Bond Polymer Refine Detection (Ref # DS98000, Leica Biosystems) with diaminobenzidine as chromogen for the visualization of the staining. The slides were counterstained with haematoxylin (Leica Biosystems). Formalin-fixed, paraffin-embedded (FFPE) human breast and colon adenocarcinomas were used as positive controls. Non-primary antibody control was also used as an additional control. All TMA slides were stained for YAP1 at the same time with the controls and cell line samples. The stained slides were scanned in an Aperio AT Turbo digital pathology system (Leica Biosystems) for visualization and evaluation. Immunohistochemistry quality control and scoring were performed by two pathologists (P.V. and J.R.C.). The immunohistochemistry scoring system employed was *H*-score, which evaluated intensity (0 to 3) and percentage of positive tumour cells (0 to 100), with a final scoring ranging from 0 to 300.

The tissue microarrays used in this study comprised 37 surgically resected lung adenocarcinoma tumours. All specimens were collected from the lung cancer tissue bank at The University of Texas M. D. Anderson Cancer Center, which is approved by the M. D. Anderson Institutional Review Board. After histological examination, tissue microarrays were constructed using three 1-mm-diameter cores per tumour. Tissue microarrays were prepared with a semi-automatic tissue arrayer (Veridiam Tissue Arrayer Model VTA-100, Veridiam) using 1 mm diameter cores in triplicate for tumours, as described previously[34]. Histological sections that were 4 μm in thickness were then prepared for the subsequent immunohistochemical analysis. Clinical and pathological information was obtained for all patients (Supplementary Table 7). Pathological tumour-node-metastasis stage had been determined for lung cancers at the time of primary tumour surgery.

Whole-exome deep sequencing and targeted sanger sequencing. Genomic DNA from NSCLC lines and patient-matched B-cell lines was isolated using DNeasy Blood and Tissue Kits (QIAGEN). Exonic DNA was captured using SureSelect 38MB All Exon Kit (Agilent) following manufacturer's protocol and sequenced using HySeq 2000 (Illumina) using a paired-end sequencing protocol with reads aligned to the NCBI human genome by Bowtie 0.12.5 as previously described[7,35], allowing for up to 2 mismatches per read. Single-nucleotide variants (SNV) were discovered from within the uniquely aligned reads, with at least one mismatch with a Phred quality score greater than 20 and coverage greater than 6 by non-redundant reads. Somatic SNVs were identified by requiring coverage on the variant site by the wild-type allele to be greater than 6 and to be 0 by the mutant allele. A series of filters was used to screen out probable germline mutations from somatic mutations identified in NSCLC cell lines without a matched normal B-cell line. They are as follows: (i) germline variants that were found in the matched dataset were removed; (ii) variants that were found to be present in dbSNP (http://www.ncbi.nlm.nih.gov/SNP/) but not in COSMIC (http://cancer.sanger.ac.uk/cosmic) were removed; (iii) silent, intergenic and untranslated region variants were removed; (iv) variants that were found at a frequency ≥8% in the thousand-genome project were removed; (v) variants in genes that were mutated ≥62 times at any site across the panel were removed; and (vi) variants that were mutated at the same amino acid position in more than 9 cell lines were removed (variants that were found to be 'hotspots' in the matched dataset were, however, rescued).

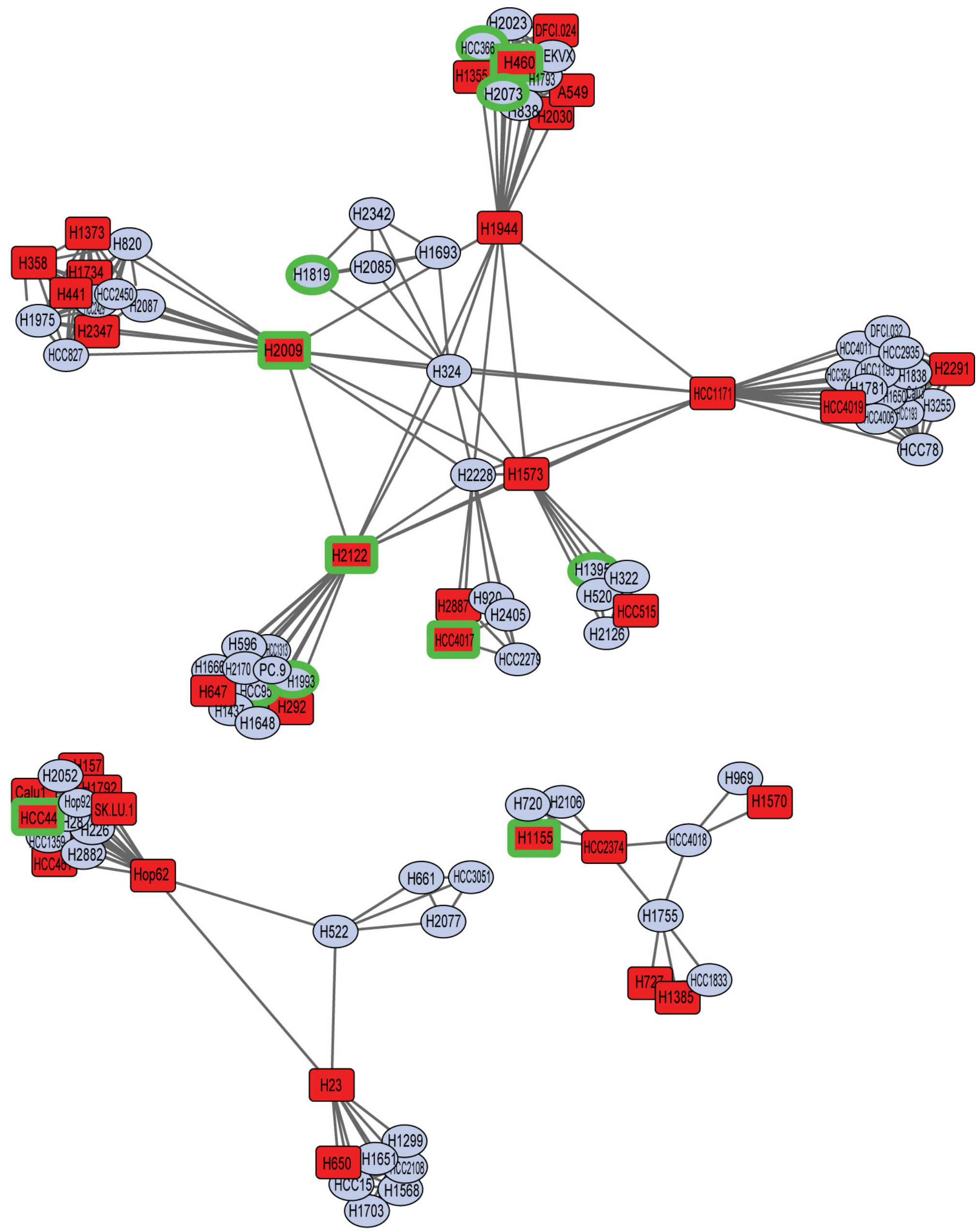
For targeted detection of mutations in *FSTL5*, genomic DNA was extracted from patient-derived human lung tumour samples and from matched normal tissue. Exons 2–16 were amplified by PCR with HotStarTqa Master Mix kit (Qiagen), purified using USB ExoSAP-IT (Affymetrix) and Sanger-sequenced. For exons that exhibited mixed chromatograms, PCR products were cloned using TOPO TA cloning kit (Invitrogen) and the resulting clones were individually sequenced.

Bicluster analysis. We converted the mutation table to a binary-presence call table in which '1' indicated the presence of a mutation in a cell line and '0' indicated the wild type. We created a biclustering script that searches for every possible permutation of rows (that is, genes) and columns (cell lines) to identify the biggest blocks of 1s in the dataset. In other words, we search for the largest number of mutations that are shared by the largest number of cell lines. We then identified the bicluster that identified mutations that were shared by *KRAS*-mutant/XPO1-inhibitor-resistant lines that were not present in the *KRAS*-mutant/XPO1-inhibitor-sensitive lines. This bicluster contained the single gene *FSTL5*.

Retrospective analysis of shRNA enrichment in HCC. Extended Data Figure 5e was generated from an oncogenomics-based *in vivo* RNAi screening result[19] (Supplementary Table 3). Shown are 36 shRNAs enriched at least 2.5-fold over the predicted representation during HCC tumour development.

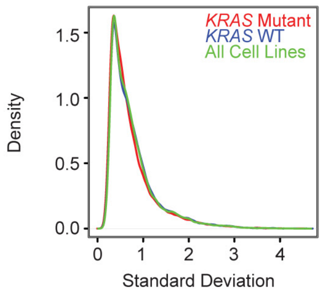
26. Singh, N. K., Seo, B. Y., Vidyasagar, M., White, M. A. & Kim, H. S. siMacro: a fast and easy data processing tool for cell-based genomewide siRNA screens. *Genomics Inform.* **11,** 55–57 (2013).
27. Shigematsu, H. *et al.* Somatic mutations of the HER2 kinase domain in lung adenocarcinomas. *Cancer Res*. **65,** 1642–1646 (2005).
28. Yamamoto, H. *et al. PIK3CA* mutations and copy number gains in human lung cancers. *Cancer Res*. **68,** 6913–6921 (2008).
29. Phelps, R. M. *et al.* NCI-Navy Medical Oncology Branch cell line data base. *J. Cell. Biochem. Suppl*. **24,** 32–91 (1996).
30. Shigematsu, H. *et al.* Clinical and biological features associated with epidermal growth factor receptor gene mutations in lung cancers. *J. Natl. Cancer Inst.* **97,** 339–346 (2005).
31. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. **13,** 2498–2504 (2003).
32. Bruckova, L. *et al.* Proliferative potential and phenotypic analysis of long-term cultivated human granulosa cells initiated by addition of follicular fluid. *J. Assist. Reprod. Genet.* **28,** 939–950 (2011).
33. DuPage, M., Dooley, A. L. & Jacks, T. Conditional mouse lung cancer models using adenoviral or lentiviral delivery of Cre recombinase. *Nat. Protocols* **4,** 1064–1072 (2009).
34. Sun, M. *et al.* HER family receptor abnormalities in lung cancer brain metastases and corresponding primary tumors. *Clin. Cancer Res.* **15,** 4829–4837 (2009).
35. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*. **10,** R25 (2009).

a

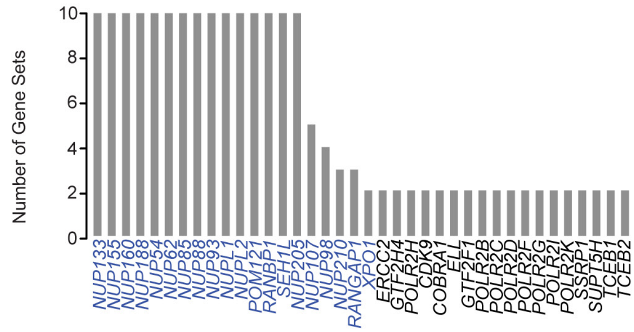


Extended Data Figure 1 | Two-dimensional APC projection of 106 NSCLC lines based on whole-genome mRNA expression variation. High-resolution, annotated version of Fig. 1a.

a



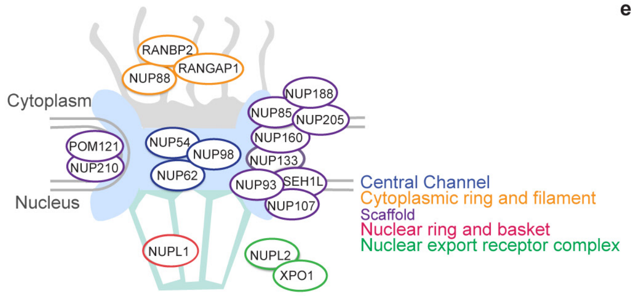
c



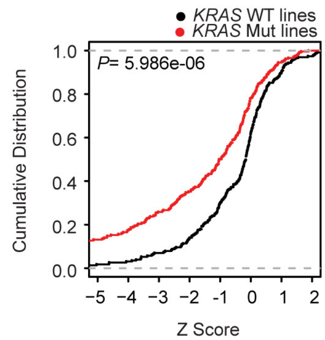
b

	Gene Set	NES	*P*-value	FDR
1	REACTOME_NEP_NS2_INTERACTS_WITH_THE_CELLULAR_EXPORT_MACHINERY	1.88	0	0.168
2	REACTOME_REV_MEDIATED_NUCLEAR_EXPORT_OF_HIV1_RNA	1.87	0	0.092
3	REACTOME_TRANSPORT_OF_THE_SLBP_INDEPENDENT_MATURE_MRNA	1.86	0	0.073
4	REACTOME_NUCLEAR_IMPORT_OF_REV_PROTEIN	1.85	0	0.063
5	REACTOME_VPR_MEDIATED_NUCLEAR_IMPORT_OF_PICS	1.83	0	0.076
6	REACTOME_TRANSPORT_OF_RIBONUCLEOPROTEINS_INTO_THE_HOST_NUCLEUS	1.82	0	0.076
7	REACTOME_REGULATION_OF_GLUCOKINASE_BY_GLUCOKINASE_REGULATORY_PROTEIN	1.78	0	0.102
8	REACTOME_LATE_PHASE_OF_HIV_LIFE_CYCLE	1.78	0	0.092
9	REACTOME_GLUCOSE_TRANSPORT	1.75	0	0.122
10	REACTOME_HIV_LIFE_CYCLE	1.70	0	0.187

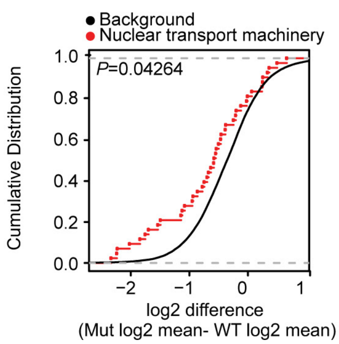
d



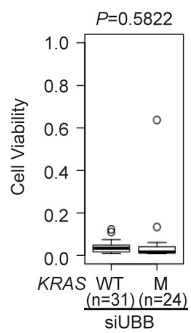
e



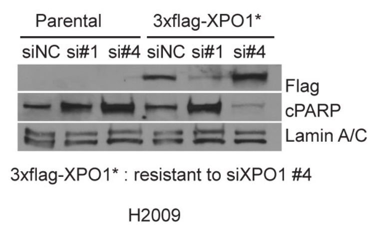
f



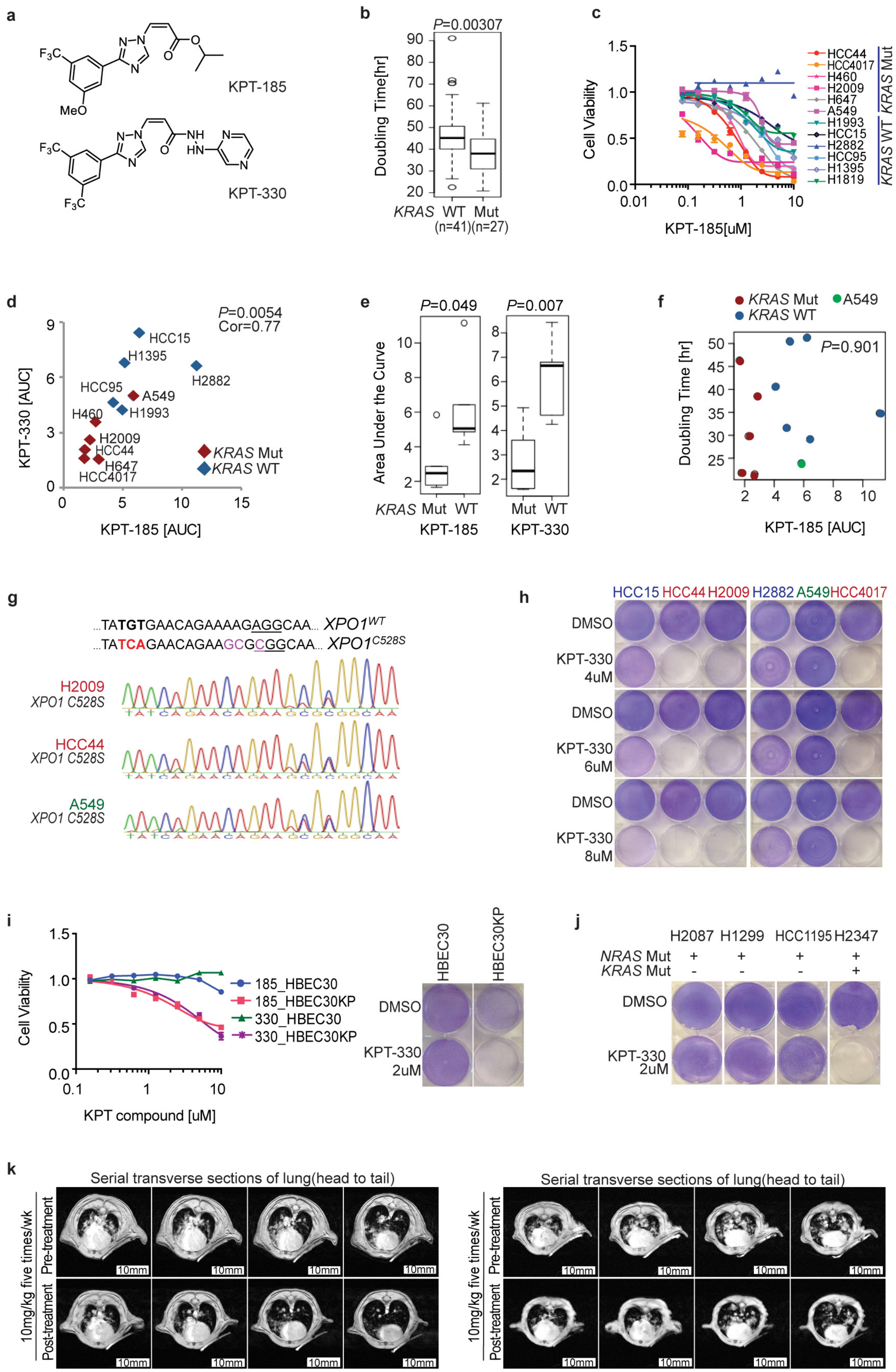
g



h

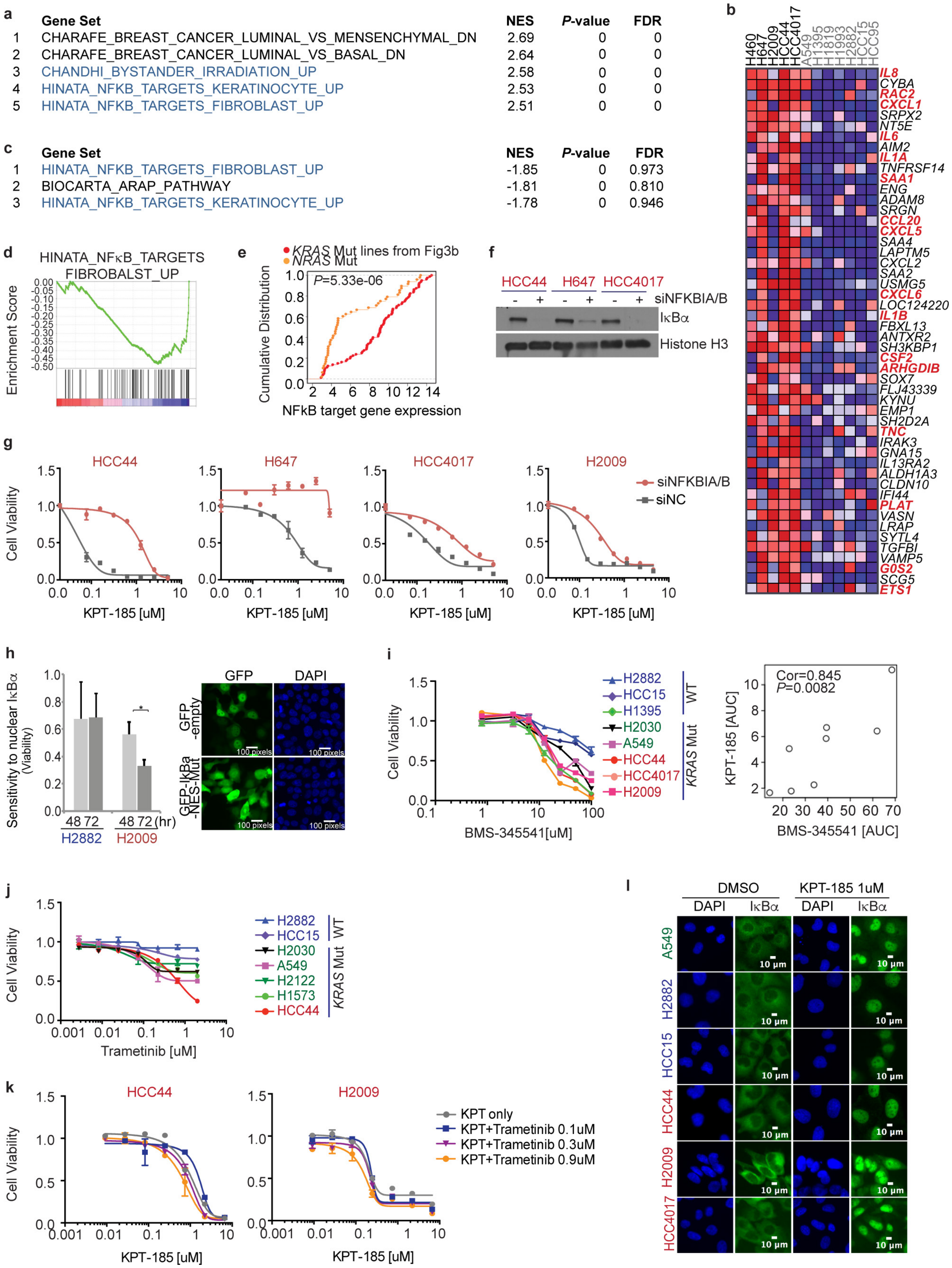


Extended Data Figure 2 | Synthetic-lethal genetic interactions in *KRAS*-mutant NSCLC cells. a, Distribution of the variation in mRNA expression among *KRAS*-mutant lines (red curve, $n = 37$), *KRAS*-wild-type lines (blue curve, $n = 69$) and all NSCLC lines (green curve). **b**, Top-ranked gene sets (FDR < 0.2, Kolmogorov–Smirnov $P < 1 \times 10^{-16}$) returned by functional GSEA. **c**, Genes present in leading edge gene representation among the gene sets in **b**. Known components of nuclear transport machinery (indicated in **d**) are labelled in blue. **d**, Biological process representation of the leading-edge synthetic-lethal gene-depletion targets. **e**, Cumulative distributions of the viability *Z*-scores for siRNA pools, targeting genes in **d**, among *KRAS*-mutant versus *KRAS*-wild-type cell lines. Kolmogorov–Smirnov test *P* value is indicated. **f**, Cumulative distributions of $\log_2$ difference scores for depletion of shRNAs in *KRAS*-mutant versus *KRAS*-wild-type cells. Red, shRNAs targeting genes encoding nuclear transport machinery from **e**; black, all other shRNAs. Kolmogorov–Smirnov test *P* value is indicated. Data obtained from a previous study[3]. **g**, Cell viability after *UBB* depletion (a broadly toxic siRNA target that serves as a positive transfection control) in *KRAS*-mutant versus wild-type lines. Unpaired *t*-test was used for the comparison. **h**, Rescue of *XPO1* siRNA toxicity by ectopic expression of a mutant *XPO1* cDNA designed to be resistant to *XPO1* siRNA number 4. Lamin A/C is shown as a loading control.



Extended Data Figure 3 | See next page for caption.

Extended Data Figure 3 | Selective sensitivity of *KRAS*-mutant NSCLC cells to chemical inhibition of the nuclear transport receptor XPO1. a, Structures of SINE compounds (XPO1 inhibitors), KPT-185 and KPT-330. **b**, Enrichment of short doubling times in *KRAS*-mutant versus wild-type NSCLC lines. Box plots indicate median and IQR. Unpaired *t*-test was used for the comparison. **c**, The 8-point dose–response viability curves for the indicated panel of NSCLC lines following a 72-h exposure to KPT-185. Mean ± s.d. ($n = 3$) is shown. **d**, Correlation of sensitivity to the XPO1 inhibitors KPT-330 and KPT-185. AUCs from **c** and Fig. 2a. Red, *KRAS* mutant; blue, *KRAS* wild type. Pearson correlation *P* value is indicated. **e**, Response of *KRAS*-mutant versus *KRAS*-wild-type cohorts to XPO1 inhibitors. Box plots indicate median values and IQR, an unpaired *t*-test was used for the comparison. **f**, Scatter plot of cell-line doubling time versus KPT-185 sensitivity. Pearson correlation *P* value is indicated. **g**, Sequencing chromatogram of *XPO1* genomic DNA of genome-edited cells. The C528S substitution was induced by CRISPR/Cas9-induced homologous recombination. Three synonymous mutations were simultaneously introduced near the PAM site (underlined) in order to prevent re-cutting of the recombined DNA. **h**, Selective sensitivity of *KRAS*-mutant lines to KPT-330 at doses over 400% higher than bioactive *in vivo* concentrations. Post-confluent cells were exposed to KPT-330 for 5 days. **i**, Response of $KRAS^{G12V}$-expressing lung epithelia (HBEC30KP), versus wild-type parental epithelia (HBEC30), to KPT-185 and KPT-330. Left, mean ± s.d., $n = 3$; right, monolayer assay is as in **h**. **j**, Cytotoxic effect of 2 μM KPT-330 on the indicated *NRAS*-mutant cell lines. Monolayer assay is as in **h**. **k**, Lung tumour burden pre- and post-treatment as indicated by magnetic resonance images. Two mice presenting with exceptionally high initial tumour burden were treated with 10 mg kg^{-1} KPT-330 five times per week. Lungs were imaged with serial transverse magnetic resonance sections on treatment day 0 and again on day 21.



a

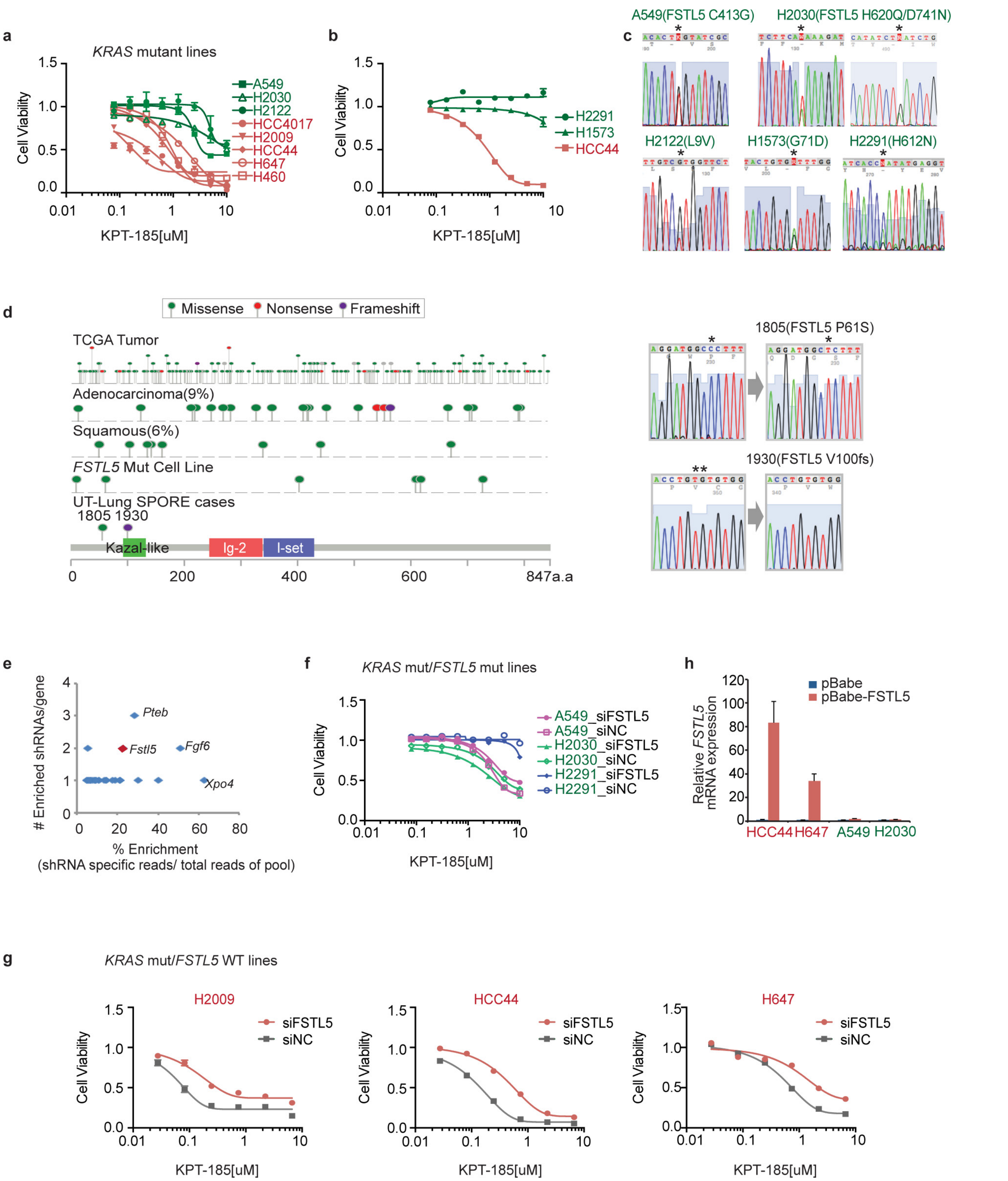
	Gene Set	NES	*P*-value	FDR
1	CHARAFE_BREAST_CANCER_LUMINAL_VS_MENSENCHYMAL_DN	2.69	0	0
2	CHARAFE_BREAST_CANCER_LUMINAL_VS_BASAL_DN	2.64	0	0
3	CHANDHI_BYSTANDER_IRRADIATION_UP	2.58	0	0
4	HINATA_NFKB_TARGETS_KERATINOCYTE_UP	2.53	0	0
5	HINATA_NFKB_TARGETS_FIBROBLAST_UP	2.51	0	0

c

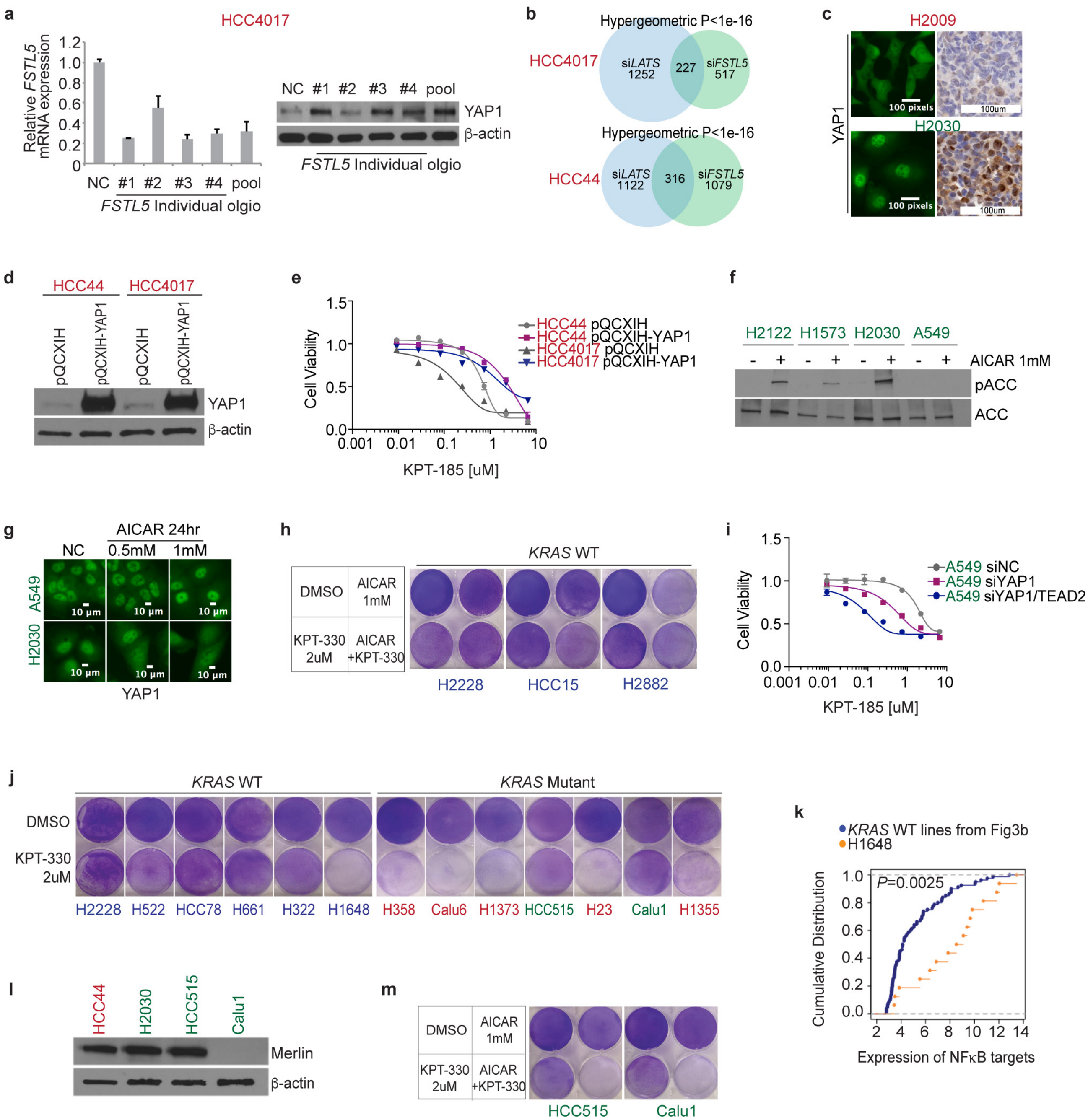
	Gene Set	NES	*P*-value	FDR
1	HINATA_NFKB_TARGETS_FIBROBLAST_UP	-1.85	0	0.973
2	BIOCARTA_ARAP_PATHWAY	-1.81	0	0.810
3	HINATA_NFKB_TARGETS_KERATINOCYTE_UP	-1.78	0	0.946

Extended Data Figure 4 | See next page for caption.

Extended Data Figure 4 | Selective addiction to NFκB activity specifies sensitivity to XPO1 inhibition. **a**, Top five gene sets that significantly discriminate XPO1-inhibitor-sensitive lines from XPO1-inhibitor-resistant lines. NFκB target gene sets are indicated in blue. **b**, Top 50 differentially expressed genes ranked by signal-to-noise (S2N) ratio. Known NFκB targets are indicated in red (16/133; hypergeometric $P < 1 \times 10^{-16}$). **c**, Top 3 gene sets that are downregulated by a 12-h exposure to an XPO1 inhibitor. NFκB target gene sets are indicated in blue. **d**, Enrichment plot of NFκB target genes (KPT-185-treated versus DMSO-treated). **e**, Evidence for attenuated NFκB signalling in XPO1-inhibitor-resistant *NRAS*-mutant cell lines. Empirical cumulative distributions of NFκB target gene expression (from **b**) are shown for *NRAS*-mutant cell lines versus *KRAS*-mutant cell lines as indicated. Yellow, *NRAS*-mutant/XPO1-inhibitor-resistant lines (H2087, H1299 and HCC1195, shown in Extended Data Fig. 3j); red, *KRAS*-mutant/XPO1-inhibitor-sensitive line (shown in Figs 2, 3). Yellow versus red, $P < 0.01$ using Kolmogorov–Smirnov test. **f**, Immunoblot of IκBα 48 h post-transfection of siNFKBIA/siNFKBIB (targeting genes that express IκBα/IκBβ) for confirmation of target depletion. Histone H3 is shown as a loading control. **g**, IκB-dependent sensitivity to KPT-185. Cells were exposed to the indicated concentrations of XPO1 inhibitors for 72 h 24 h post-transfection with the indicated siRNA. Mean and range ($n = 2$). **h**, Intolerance to ectopic nuclear accumulation of IκBα in XPO1-inhibitor-sensitive cells. Left, *y* axis indicates fold change in the percentage of GFP-positive nuclei of GFP-IκB-NES-mutant-positive cells normalized to GFP-empty-vector-positive cells. Bars indicate mean ± s.d. for three independent experiments (*$P < 0.05$, Unpaired *t*-test). Right, 293T cells transfected with the indicated plasmids to confirm plasmid transfection efficiency and localization of ectopically expressed proteins. Cells were fixed and photographed 48 h post-transfection. **i**, Positive correlation between sensitivity to KPT-185 and BMS-345541 ($P < 0.01$, Pearson correlation). Dose–response curves of a panel of NSCLC lines following a 72-h exposure to BMS-345541. Mean ± s.d. ($n = 3$). AUCs of KPT-185 were determined from Extended Data Fig. 3c. Red labels, *KRAS* mutant/XPO1-inhibitor sensitive; green labels, *KRAS* mutant/XPO1-inhibitor resistant; blue labels, *KRAS* wild type. **j**, Dose–response curves of a panel of NSCLC lines following a 72-h exposure to Trametinib. Mean and range ($n = 2$). Label colours as in **i**. **k**, Dose–response curves of a panel of NSCLC lines following a 72-h exposure to KPT-185 combined with the indicated concentrations of Trametinib. Mean and range ($n = 2$). **l**, Subcellular localization of IκBα in the presence of 1 μM KPT-185. Cells were exposed to KPT-185 for 24 h. Label colours are as in **i**.



Extended Data Figure 5 | Concurrent mutations in *FSTL5* are associated with intrinsic resistance of *KRAS*-mutant lines to XPO1 inhibitor. a, Eight-point dose–response viability curves for H2122 and H2030 following a 72-h exposure to KPT-185. Mean ± s.d. ($n = 3$). Data are overlaid with responses of the indicated lines from Extended Data Fig. 3c for comparison. **b**, KPT-185 dose–response curves. Mean ± s.d. ($n = 3$). **c**, *FSTL5* Sanger-sequencing chromatograms of detected *FSTL5* variants in the indicated cell lines. **d**, Map of somatic alterations in *FSTL5* detected in all cancers (TCGA), lung adenocarcinoma (TCGA), lung squamous (TCGA), NSCLC cell lines (this study), and human lung tumour samples, 1805 and 1930 (this study). **e**, Tumour suppressor genes identified in an oncogenomics-based *in vivo* RNAi screen[19]. Among the genes targeted by 36 shRNAs overrepresented during HCC tumour development, *Fstl5* was the third ranked gene suppressed by >1 enriched shRNA. The *y* axis indicates number of shRNAs per gene among the 36 enriched shRNAs. The *x* axis indicates shRNA specific reads over a total 2,307 sequence reads. **f**, **g**, KPT-185 dose–response of cells transfected with the indicated siRNAs as in Extended Data Fig. 4g. **f** shows *KRAS*-mutant/*FSTL5*-wild-type lines, **g** shows *KRAS*-mutant/*FSTL5* mutant lines. Mean and range ($n = 2$). **h**, Relative ectopic expression of *FSTL5* mRNA. Cells were infected with retrovirus carrying the indicated plasmids. Following a 7-day puromycin selection, cells were collected for qPCR. Mean and range ($n = 2$).



Extended Data Figure 6 | See next page for caption.

Extended Data Figure 6 | Concurrent mutations in *FSTL5* are mechanistically coupled to YAP1 activation. a, Expression of *FSTL5* mRNA (left) and YAP1 protein (right) following transfection with the indicated siRNAs targeting *FSTL5*. Cells were collected 72 h post-transfection for parallel qPCR and immunoblotting. Mean and range ($n = 2$). **b**, Intersection of the FSTL5-dependent and LATS-dependent gene expression programs in *KRAS*-mutant/XPO1-inhibitor-sensitive NSCLC lines. To evaluate the enrichment of YAP-responsive genes within the FSTL5-dependent gene expression network, quantitative whole-genome transcript arrays were prepared with mRNA isolated from the indicated cell lines treated with the indicated siRNAs 72 h post-transfection. *LATS1/2* depletion was used to activate YAP-dependent gene expression. All arrays were normalized to corresponding control siRNA-treated samples. Euler plots indicate genes up- or downregulated at least twofold in response to siFSTL5, siLATS or both; hypergeometric P values are indicated. **c**, YAP1 fluorescence micrographs and representative YAP1 immunohistochemistry. H2009 and H2030 cell lines were used as a negative and positive control for YAP1 staining, respectively. **d**, Stably overexpressed YAP1 in *KRAS*-mutant/XPO1-inhibitor-sensitive lines. Cells were infected with the indicated retroviral vector, selected with hygromycin and then collected for immunoblotting. **e**, Induction of XPO1-inhibitor resistance by YAP1 overexpression. Proliferating cells stably expressing indicated plasmids were exposed to XPO1 inhibitors for 3 days. Mean ± range ($n = 2$). **f**, Immunoblot of the indicated proteins in *KRAS*-mutant/XPO1-inhibitor-resistant lines following a 24-h exposure to 1 mM AICAR. AICAR resulted in accumulation of phospho-acetyl-CoA-carboxlyase (pACC), an indicator of AMPK activation in all the lines tested except A549. A549 is known to be non-responsive to AICAR owing to the absence of LKB1 (also known as STK11) expression[7]. **g**, Subcellular localization of YAP1 in response to 0.5 or 1 mM AICAR. Cells were exposed to AICAR for 24 h. Cytoplasmic accumulation of YAP1 was observed in response to AICAR exposure in H2030, but not in A549. **h**, Resistance of *KRAS*-wild-type lines to KPT-330 in combination with AICAR. Post-confluent cells were exposed to the indicated compounds for 3 days. **i**, Induction of XPO1 inhibitor-sensitivity by *YAP1* and *TEAD2* depletion. 48 h post-transfection with the indicated siRNAs, cells were exposed to the indicated concentrations of XPO1 inhibitors for 3 days. Mean and range ($n = 2$). **j**, Cytotoxic effect of 2 μM KPT-330 on indicated cell lines. Post-confluent cells were exposed to KPT-330 for 5 days. Red labels, *KRAS*-mutant/XPO1-inhibitor-sensitive; green labels, *KRAS*-mutant/XPO1-inhibitor-resistant; blue labels, *KRAS* wild type. **k**, Evidence for NFκB pathway activation in H1648 cells. Empirical cumulative distributions of NFκB target gene expression (from Extended Data Fig. 4b) are shown for H1648 versus *KRAS*-wild-type cell lines as indicated. Blue, *KRAS*-wild-type/XPO1-inhibitor-resistant lines (H2882, HCC15, H1395, H1993 and HCC95 shown in Figs 2 and 3); yellow, *KRAS*-wild-type/XPO1-inhibitor-sensitive line H1648 (shown in **j**). Cancer Cell Line Encyclopedia data indicates that H1648 harbours genomic amplification of *IKBKB*. Blue versus yellow, $P < 0.01$, Kolmogorov–Smirnov test. **l**, Merlin expression is absent in Calu1 cells. **m**, Cytotoxic effect of the indicated compounds on the indicated cell lines. Post-confluent cells were treated as in **h**. HCC515 harbours a somatic mutation in *LATS1*.

doi:10.1038/nature19759

Acetylation-regulated interaction between p53 and SET reveals a widespread regulatory mode

Donglai Wang[1]*, Ning Kon[1]*, Gorka Lasso[2], Le Jiang[1], Wenchuan Leng[3], Wei-Guo Zhu[4], Jun Qin[3,5], Barry Honig[2] & Wei Gu[1]

Although lysine acetylation is now recognized as a general protein modification for both histones and non-histone proteins[1–3], the mechanisms of acetylation-mediated actions are not completely understood. Acetylation of the C-terminal domain (CTD) of p53 (also known as TP53) was an early example of non-histone protein acetylation[4] and its precise role remains unclear. Lysine acetylation often creates binding sites for bromodomain-containing 'reader' proteins[5,6]. Here we use a proteomic screen to identify the oncoprotein SET as a major cellular factor whose binding with p53 is dependent on CTD acetylation status. SET profoundly inhibits p53 transcriptional activity in unstressed cells, but SET-mediated repression is abolished by stress-induced acetylation of p53 CTD. Moreover, loss of the interaction with SET activates p53, resulting in tumour regression in mouse xenograft models. Notably, the acidic domain of SET acts as a 'reader' for the unacetylated CTD of p53 and this mechanism of acetylation-dependent regulation is widespread in nature. For example, acetylation of p53 also modulates its interactions with similar acidic domains found in other p53 regulators including VPRBP (also known as DCAF1), DAXX and PELP1 (refs. 7–9), and computational analysis of the proteome has identified numerous proteins with the potential to serve as acidic domain readers and lysine-rich ligands. Unlike bromodomain readers, which preferentially bind the acetylated forms of their cognate ligands, the acidic domain readers specifically recognize the unacetylated forms of their ligands. Finally, the acetylation-dependent regulation of p53 was further validated *in vivo* by using a knock-in mouse model expressing an acetylation-mimicking form of p53. These results reveal that acidic-domain-containing factors act as a class of acetylation-dependent regulators by targeting p53 and, potentially, other proteins.

Although the physiological consequences of acetylation at positions K120 and K164 within the DNA-binding domain have been established in studies of p53 acetylation-defective mutant mice[10,11], the *in vivo* functions of CTD acetylation remain unclear. By examining mutant mice expressing C-terminal truncated forms of p53, two recent studies have shown that loss of the CTD results in p53 activation[12,13], suggesting that the CTD may act as a docking site for negative regulators of p53. Nevertheless, the identity of the negative regulators and the consequences of CTD acetylation remain unknown. To identify proteins that bind to p53 in a manner dependent on the CTD acetylation status of p53, we synthesized both unacetylated (Un-Ac) and fully-acetylated (Ac) biotin-conjugated CTD peptides and used the immobilized peptides as affinity columns to purify cellular factors (Fig. 1a). We failed to identify any proteins enriched in the acetylated p53 CTD column (Fig. 1b). Instead, coomassie blue staining of the bound fraction revealed a major band of approximately 38 kDa from the unacetylated p53 column that was completely absent in the acetylated column. Mass spectrometry analysis of this band revealed 28 unique peptides identical to SET (Fig. 1c and Extended Data Fig. 1a), an oncoprotein that is activated by translocation-associated gene fusions in patients with acute myeloid leukaemia[14]. Although a previous study reported an interaction between p53 and SET[15], the impact of CTD acetylation on the functional consequences of this interaction are unclear.

Acetylation-dependent disruption of the p53–SET interaction was confirmed *in vitro* with purified SET protein (Fig. 1d). Moreover, expression of CREB-binding protein (CBP), the enzyme responsible for CTD acetylation, completely abrogated the formation of SET complexes with wild-type p53 ($p53^{WT}$), but not with a CTD acetylation-deficient p53 ($p53^{KR}$) mutant, confirming that CTD acetylation is crucial for the p53–SET interaction in cells (Fig. 1e). Notably, other modifications on the CTD lysine residues, including methylation, ubiquitination, sumoylation and neddylation, had no effect on this binding, underscoring the specificity of the acetylation-dependent control of p53–SET interactions (Extended Data Fig. 1b–e).

Next, we tested whether SET acts as a transcriptional cofactor by forming a p53–SET complex on the p53 target promoter. Although SET alone showed no obvious DNA-binding activity (Fig. 1f), in the presence of both p53 and SET, a slower-migrating SET/p53–DNA complex was formed and super-shifted by antibodies against p53 or SET. Further binding-domain mapping indicated that the CTD of p53 interacts directly with the acidic domain of SET (Extended Data Fig. 1f–h). To determine the impact of SET on the transcriptional activity of p53, we measured transactivation of a p53-responsive reporter gene. Indeed, p53-mediated transactivation was abrogated upon co-expression of wild-type SET, but not a SET mutant lacking the acidic domain required for p53 binding (Fig. 1g). Conversely, wild-type SET-mediated repression was abrogated when a p53 mutant lacking the CTD was expressed (Fig. 1g). Notably, the interaction of endogenous p53 and SET was easily detected in unstressed cells; however, upon DNA damage, despite increased p53 levels, the p53–SET interaction was largely diminished, probably owing to the induction of CTD acetylation (Fig. 1h). Moreover, chromatin immunoprecipitation (ChIP) assays revealed that the recruitment of SET to the promoter of p53 targets was largely inhibited (Fig. 1i and Extended Data Fig. 1i–k). Together, these data indicate that SET acts as a transcriptional co-repressor of p53. However, acetylation of the CTD upon DNA damage leads to abrogation of this repression through disruption of the p53–SET interaction (Fig. 1j).

We further investigated whether inactivation of SET influences the activities of p53 in human cancer cells. RNA-interference-mediated depletion of SET markedly elevated the expression of p53 targets, such as cyclin dependent kinase inhibitor 1A (CDKN1A, also known as p21) and p53 upregulated modulator of apoptosis (PUMA, also

[1]Institute for Cancer Genetics, Department of Pathology and Cell Biology, Herbert Irving Comprehensive Cancer Center, College of Physicians & Surgeons, Columbia University, 1130 Nicholas Ave, New York, New York 10032, USA. [2]Department of Biochemistry and Molecular Biophysics and Systems Biology, Center for Computational Biology and Bioinformatics, Howard Hughes Medical Institute, Columbia University, 1130 Nicholas Ave, New York, New York 10032, USA. [3]State Key Laboratory of Proteomics, National Center for Protein Sciences (The PHOENIX Center, Beijing), Beijing, 102206, China. [4]Department of Biochemistry and Molecular Biology, Shenzhen University School of Medicine, Shenzhen 518060, China. [5]Alkek Center for Molecular Discovery, Verna and Marrs McLean Department of Biochemistry and Molecular Biology, Department of Molecular and Cellular Biology, Baylor College of Medicine, Houston, Texas 77030, USA.
*These authors contributed equally to this work.

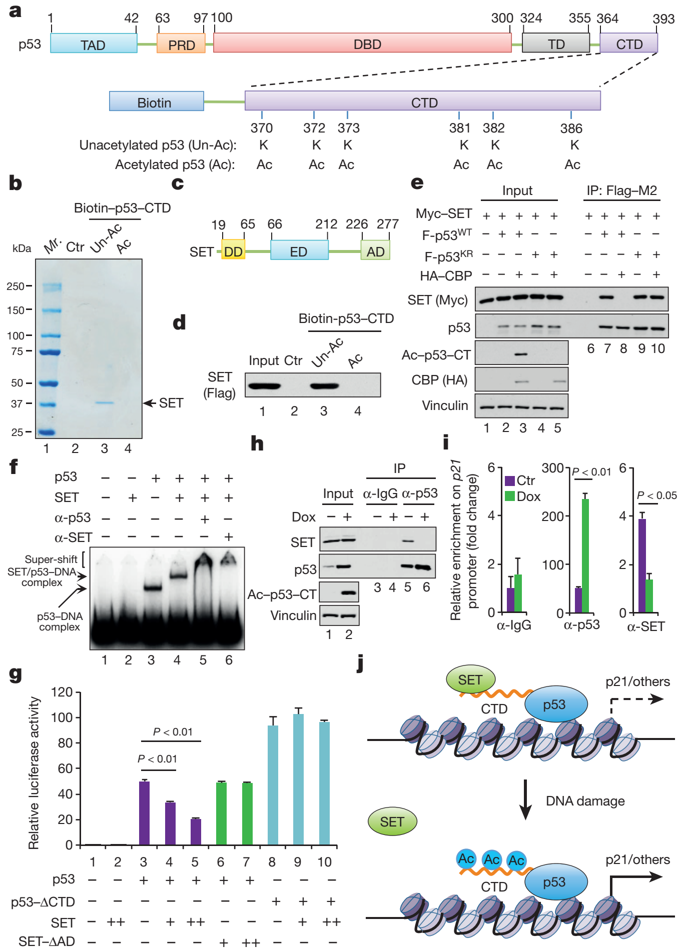


Figure 1 | Identification of SET as a specific co-repressor of C-terminal unacetylated p53. a, Schematic diagram of the synthesized biotin-conjugated p53 CTD. **b**, Coomassie blue staining of the protein complex bound with the p53 CTD. **c**, Schematic diagram of SET. DD: dimerization domain; ED: earmuff domain; AD: acidic domain. **d**, *In vitro* binding assay of p53 CTD and purified SET. **e**, Western blot analysis of the interaction between p53 and SET in the nuclear fraction of H1299 cells. **f**, Electrophoretic mobility shift assay showing the SET/p53–DNA complex formation *in vitro*. **g**, Luciferase assays of SET-mediated regulation of p53 transactivity in H1299 cells. **h**, Western blot analysis of the endogenous interaction between p53 and SET upon doxorubicin (Dox) treatment of HCT116 cells. **i**, ChIP analysis of p53 or SET recruitment onto the *p21* promoter upon Dox treatment of HCT116 cells. **j**, A model of dynamic promoter-recruitment of SET regulated by p53 CTD acetylation status. Error bars indicate mean ± s.d., $n = 3$ for technical replicates. Data are shown as representative of three experiments. Uncropped blots can be found in Supplementary Fig. 1.

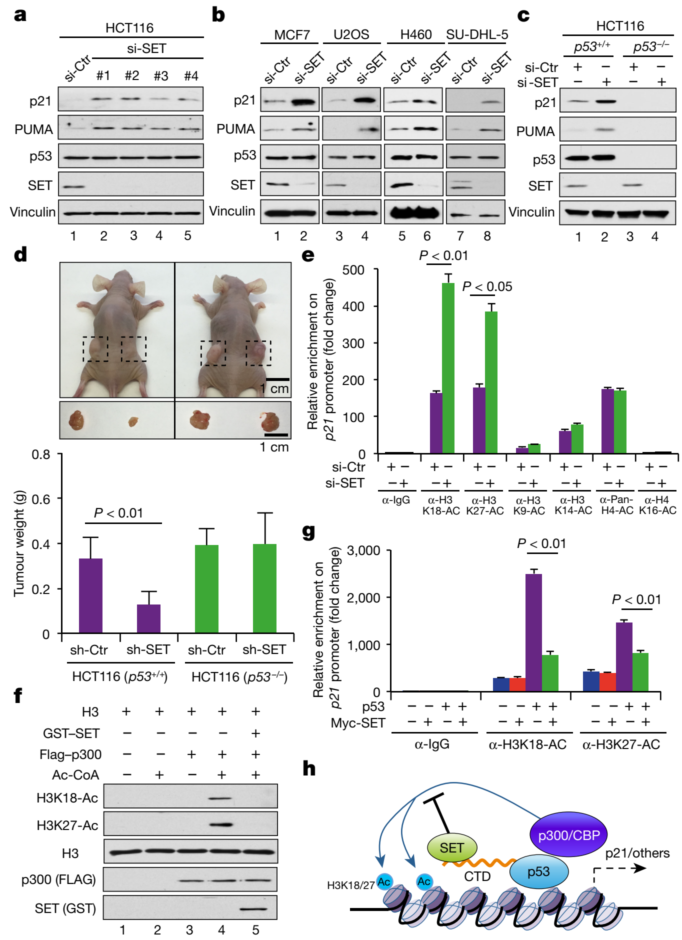


Figure 2 | SET negatively regulates p53 transactivity by inhibiting p300/CBP-mediated H3K18 and H3K27 acetylation on the p53 target promoter. a–c, Western blot analysis of the effect of SET knockdown on p53 activity in cells. si-Ctr: control siRNA. **d**, Xenograft analysis of SET-mediated effect on growth of control and p53-deficient HCT116 tumours. Top, representative images of mice (NU/NU; left flank: control knockdown cells; right flank: SET knockdown cells). Insert: images of dissected HCT116 tumours from the mice shown above. Bottom, analysis of tumour weight growing from $p53^{+/+}$ and $p53^{-/-}$ HCT116 cells after SET depletion in xenografted mice. sh-Ctr: control shRNA; sh-SET: human SET-specific shRNA. Scale bars, 1 cm. **e**, ChIP analysis of the SET knockdown-mediated effect on histone modifications at the *p21* promoter in HCT116 cells. **f**, *In vitro* acetylation assay of the effect of SET on p300-mediated H3K18 and H3K27 acetylation. **g**, ChIP analysis of the SET-mediated effect on p53-dependent H3K18 and H3K27 acetylation on the *p21* promoter in H1299 cells. **h**, A model of SET-mediated regulation on p53 transactivity. Error bars indicate mean ± s.d., $n = 3$ for technical replicates in **e** and **g**; $n = 5$ ($p53^{+/+}$ group) or $n = 3$ ($p53^{-/-}$ group) for biological replicates in **d**. Data are shown as representative of three experiments. Uncropped blots can be found in Supplementary Fig. 1.

known as Bcl-2-binding component 3), without affecting the steady-state levels of endogenous p53 in HCT116 colorectal carcinoma cells (Fig. 2a). Similar effects were obtained in other human cancer cell lines that express wild-type p53, including MCF7 (breast carcinoma), U2OS (osteosarcoma), H460 (lung carcinoma) and SU-DHL-5 (B-cell lymphoma) (Fig. 2b). Moreover, this induction of p21 and PUMA expression was completely abrogated in isogenic HCT116 $p53^{-/-}$ cells (Fig. 2c), indicating that the SET-mediated effects are p53-dependent. Further analysis of U2OS and p53-null U2OS cells that had SET knocked down identified a number of p53 targets that were upregulated upon inactivation of SET in a p53-dependent manner; SET knockdown induced p53-dependent cell growth repression in those cells (Extended Data Figs 2a–c, 3a, b). To examine the effect of SET on p53-mediated tumour suppression, we tested whether SET depletion affected cell growth in xenograft tumour models in immunodeficient mice (NU/NU). SET knockdown strongly suppressed tumour growth of HCT116 cells, but not isogenic HCT116 $p53^{-/-}$ cells (Fig. 2d). Moreover, the p53-dependent effects were further validated in HCT116 p53 knockout cells generated by the CRISPR/Cas9-mediated genome editing technique (Extended Data Fig. 3c–e). These data indicated that the p53–SET interaction is crucial for the tumour growth suppression induced by p53.

As SET had no apparent effect on protein stability, DNA binding or acetylation levels of p53 (Extended Data Fig. 4a–c), we examined whether SET suppressed p53-mediated transactivation by affecting chromatin modifications at p53 target promoters. ChIP analysis revealed that SET depletion significantly increased the acetylation levels of H3K18 and H3K27 at the promoters of *p21* and *PUMA* in

HCT116 cells without affecting H3K9, H3K14, H4K16 or inducing pan-H4 acetylation (Fig. 2e and Extended Data Fig. 4d). p300/CBP, which target H3K18 and H3K27 acetylation *in vivo*[16,17], act as a key co-activators of p53-mediated transcriptional activation[18–20]. We tested whether SET suppressed p300/CBP-mediated acetylation of H3K18 and H3K27, as SET had no obvious effect on the recruitment of p300/CBP (Extended Data Fig. 4e). Indeed, *in vitro* acetylation assays revealed that SET effectively suppressed p300-dependent acetylation of H3K18 and H3K27 (Fig. 2f) and these findings were further verified for p53 target promoters by ChIP analysis (Fig. 2g and Extended Data Fig. 4f). Together, these data indicate that SET represses p53-mediated transactivation by inhibiting p300/CBP-dependent acetylation of H3K18 and H3K27 on p53 target promoters (Fig. 2h).

Numerous studies have indicated that lysine acetylation often creates docking sites for ‘reader’ proteins that possess a bromodomain, a structural motif that forms a recognition surface for acetylated lysine[5,6]. Our analysis of the p53–SET interaction suggests that the acidic domain of SET serves as a ‘converse reader’ that binds the lysine-rich CTD of p53 in a manner that can be specifically abrogated upon acetylation of these lysine residues. To further evaluate this model, we tested whether p53 interacts with other proteins in a similar manner. Several transcription cofactors known to interact directly with p53, including VPRBP, DAXX and PELP1 (refs. 7–9), also contain acidic domains similar to that of the SET protein (Fig. 3a and Extended Data Fig. 5a). Their acidic domains also readily bound unacetylated, but not acetylated, p53 CTD (Fig. 3b–d). Similar results were also obtained when the full-length proteins of VPRBP, DAXX and PELP1 were tested (Extended Data Fig. 5b). More importantly, the interactions of VPRBP, DAXX and PELP1 with wild-type p53, but not the acetylation-deficient $p53^{KR}$ mutant, were inhibited by CBP-induced acetylation in human cells (Extended Data Fig. 5c–e).

Previous studies showed that SET also regulates the activities of several other cellular factors, including histone H3, KU70 and FOXO1, through direct interactions with these proteins[21–23]. Notably, the binding region of all three proteins contains a lysine-rich domain (KRD) similar to the CTD of p53 (Fig. 3e). These lysine residues have also been reported to be acetylated *in vivo*[24–26]. To test whether SET-mediated interactions with these factors are also regulated by acetylation, we performed *in vitro* binding assays of the acidic domain of SET with unacetylated or acetylated KRDs of H3, KU70 and FOXO1. The acidic domain of SET interacted with unacetylated, but not acetylated, KRDs of H3, KU70 and FOXO1 (Fig. 3f–h). Similar results were also obtained when the full-length SET protein was used in the binding assays (Extended Data Fig. 5f–h), suggesting that the interaction of SET with H3, KU70 and FOXO1 were abrogated by acetylation in a manner analogous to that of p53 binding to SET. Since VPRBP, DAXX and PELP1 have also been implicated in transcription regulation, we investigated whether these factors could interact with H3 in a similar manner. VPRBP, DAXX and PELP1 specifically bound unacetylated H3 whereas, as expected, bromodomain proteins such BRD4 and BRD7 recognized only acetylated H3 (Extended Data Fig. 5i, j).

Our data indicate that this mechanism of acetylation-dependent regulation is widespread in nature. As the positive charge within the KRD can attract the negative charge of the acidic domain, these lysine clusters form a docking site for acidic-domain-containing regulators. However, upon acetylation, the positive charge of the lysine sidechains is neutralized, abolishing the docking site for the acidic-domain-containing regulators. Conversely, deacetylation of these lysine residues reverses this effect and promotes the recruitment of acidic-domain-containing regulators (Fig. 3i). Thus, unlike bromodomain readers, which preferentially bind the acetylated forms of their cognate ligands, the acidic domain readers specifically recognize the unacetylated forms of their ligands.

To corroborate this notion, we compared the SET-binding properties of the acetylation-deficient mutant $p53^{KR}$ with an acetylation-mimicking mutant, $p53^{KQ}$ (Extended Data Fig. 6a)., The $p53^{KR}$ mutant,

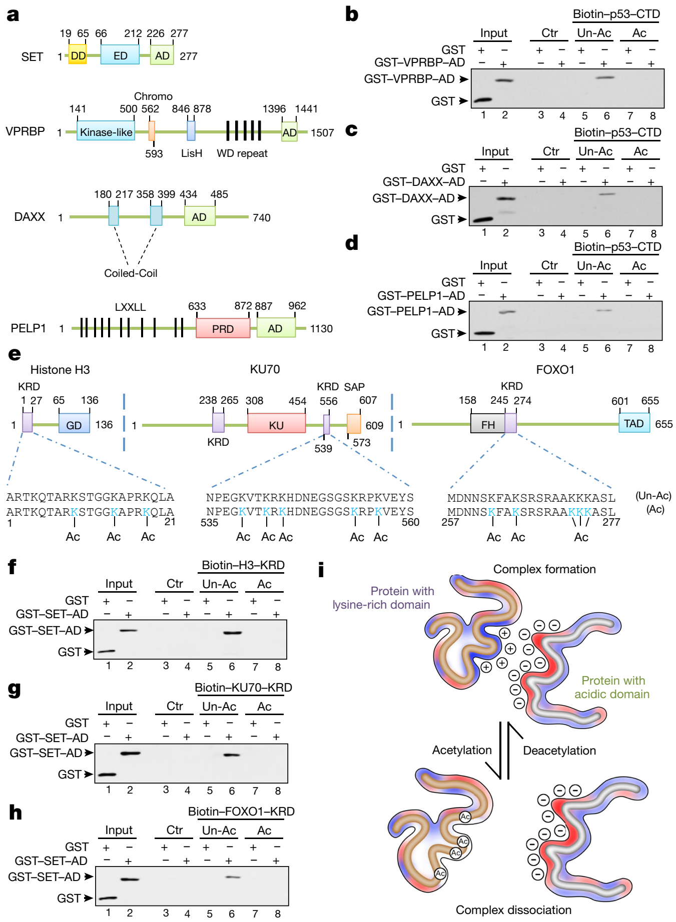


Figure 3 | Acidic-domain-containing proteins represent a new class of ‘reader’ for their unacetylated ligands. a, Schematic diagrams of the acidic-domain (AD)-containing proteins SET, VPRBP, DAXX and PELP1. **b–d**, *In vitro* binding assay of p53 CTD and acidic domains of VPRBP (**b**), DAXX (**c**) and PELP1 (**d**). Empty streptavidin beads were used as negative binding control (Ctr). **e**, Schematic diagrams of the KRD-containing proteins histone H3, KU70 and FOXO1. **f–h**, *In vitro* binding assay between the purified SET acidic domain and KRDs of H3 (**f**), KU70 (**g**) and FOXO1 (**h**). **i**, A model of acetylation-dependent regulation of the interactions between KRD-containing proteins and their acidic-domain-containing ‘readers’. Uncropped blots can be found in Supplementary Fig. 1.

like unacetylated p53, strongly bound SET (Extended Data Fig. 6b); conversely, the $p53^{KQ}$ mutant, like acetylated p53, did not interact with SET. Similar results were also obtained upon analysis of the acetylation-modulated interactions of p53 with VPRBP, DAXX and PELP1 (Extended Data Fig. 6c–e).

To further determine the physiological importance of these interactions *in vivo*, we generated $p53^{KQ/KQ}$-mutant mice (Extended Data Fig. 7a–d). Although heterozygous $p53^{+/KQ}$ mice displayed normal postnatal development, $p53^{KQ/KQ}$ homozygous mice showed neonatal lethality (Extended Data Fig. 7e). All newborn $p53^{KQ/KQ}$ pups were slightly smaller than their $p53^{+/+}$ littermates (Fig. 4a), lacked milk in their stomachs and died within one day of birth, apparently owing to dehydration from lack of maternal nourishment. In addition, live $p53^{KQ/KQ}$ mice also displayed uncoordinated movements, consistent with neurological impairments. Indeed, the brains of $p53^{KQ/KQ}$ mice appeared smaller than those of $p53^{+/+}$ mice (Fig. 4b).

Immunohistochemistry analysis of $p53^{KQ/KQ}$ brain sections revealed a marked induction of cleaved caspase 3 staining without an obvious increase in p53 protein levels (Fig. 4c and Extended Data Fig. 7f), suggesting that the neurological defects of $p53^{KQ/KQ}$ mice may reflect

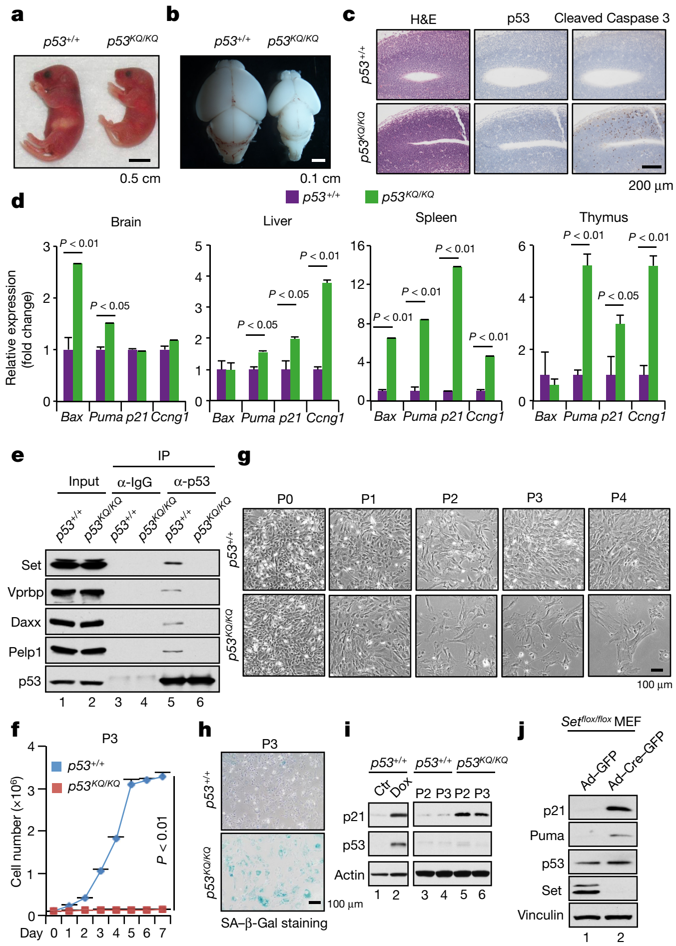


Figure 4 | The physiological significance of acetylation-dependent dissociation of p53 from its acidic-domain-containing 'readers'. **a**, Newborn $p53^{+/+}$ and $p53^{KQ/KQ}$ mice. Scale bar, 0.5 cm. **b**, The brains of newborn $p53^{+/+}$ and $p53^{KQ/KQ}$ mice. Scale bar, 0.1 cm. **c**, Immunohistochemistry analysis of brain sections from $p53^{+/+}$ and $p53^{KQ/KQ}$ embryos. Scale bar, 200 μm. **d**, RT–qPCR analysis of gene expression of p53 targets in $p53^{+/+}$ and $p53^{KQ/KQ}$ tissues. **e**, Western blot analysis of the interaction between p53 and acidic-domain-containing proteins in $p53^{+/+}$ or $p53^{KQ/KQ}$ MEFs treated with the proteasome inhibitor epoxomicin. **f**, Cell growth analysis of $p53^{+/+}$ or $p53^{KQ/KQ}$ MEFs at passage 3 (P3). **g**, Morphological representative images of $p53^{+/+}$ and $p53^{KQ/KQ}$ MEFs from P0 to P4. Scale bar, 100 μm. **h**, SA–β-gal staining of $p53^{+/+}$ and $p53^{KQ/KQ}$ MEFs (P3). Scale bar, 100 μm. **i**, Western blot analysis of p21 and p53 expression in $p53^{+/+}$ and $p53^{KQ/KQ}$ MEFs. **j**, Western blot analysis of p53 targets in *Set* conditional knockout MEFs. Error bars indicate mean ± s.d., $n = 3$ for technical replicates in **d**; $n = 3$ for biological replicates in **f**. Data are shown as representative of three experiments. Uncropped blots can be found in Supplementary Fig. 1.

increased apoptosis due to deregulation of the $p53^{KQ}$ protein. In accordance with this notion, the major apoptotic transcriptional targets of p53, namely *Bax* and *Puma*, were significantly upregulated in $p53^{KQ/KQ}$ brain tissue (Fig. 4d). Indeed, various tissues of $p53^{KQ/KQ}$ mice displayed distinct patterns of induction of different p53 target genes, suggesting tissue-specific activation of target genes by $p53^{KQ}$ *in vivo* (Fig. 4d).

The p53–SET interaction was readily detected in $p53^{+/+}$, but not $p53^{KQ/KQ}$, mouse embryonic fibroblasts (MEFs) (Fig. 4e). Similar results were also obtained for the other acidic-domain-containing cofactors (VPRBP, DAXX and PELP1), suggesting that the $p53^{KQ}$ mutant recapitulates the activity of acetylated p53 *in vivo*. Moreover, $p53^{KQ/KQ}$ MEFs displayed a severe proliferation defect (Fig. 4f) and exhibited clear signs of senescence, including a flat and enlarged morphology with large multinucleated nuclei and marked senescence-associated β-galactosidase (SA-β-Gal) staining (Fig. 4g, h and Extended Data Fig. 7g, h). In addition, western blot analysis revealed an increase in the steady-state levels of p21 protein in $p53^{KQ/KQ}$ MEFs (Fig. 4i). To directly address the role of SET *in vivo*, we generated *Set*-mutant mice (Extended Data Fig. 8a, b). Although the characterization of these mice was not complete (Extended Data Fig. 8c–e), we prepared $Set^{flox/flox}$ MEFs for functional analysis. As shown in Fig. 4j, upon Cre-mediated *Set* deletion, the expression of p53 target genes, such as p21 and Puma, was markedly induced, indicating that SET is a critical regulator of p53 *in vivo*. Together, these data validate the key role of CTD acetylation in p53 activation *in vivo*.

Previous studies showed that a $p53^{KR}$ knock-in mutant targeting the same CTD lysine residues does not significantly affect mouse development or p53 activity in mouse tissues or embryonic fibroblasts[27,28]. Thus, loss of modifiable CTD lysines may neutralize the overall effect on p53 function by abrogating both the negative and positive effects of regulation through different types of CTD modification. Surprisingly, $p53^{KQ}$ knock-in mice died shortly after birth with substantial p53 activation. Like $p53^{KR}$, $p53^{KQ}$ also eliminates other types of modification on these lysine residues; however, $p53^{KQ}$ mimics the acetylated form while $p53^{KR}$ resembles unacetylated p53. Thus, the difference between the phenotypes of $p53^{KQ}$ and $p53^{KR}$ mutant mice underscores the role of CTD acetylation *in vivo*.

The acidic-domain-containing proteins in this study consist of a specific group of proteins that harbour long clusters of acidic amino acids. Searching the Uniprot database with our motif-finding algorithm[29], we identified 49 polypeptides with highly acidic domains similar to SET, many of which are involved in transcriptional regulation and chromatin remodelling (Extended Data Table 1). In addition, by using the Species-Specific Prediction of lysine (K) Acetylation program (SSPKA)[30], we also identified 49 proteins containing a cluster of lysine residues that can potentially bind these acidic domains in an acetylation-modulated manner (Extended Data Table 2). On the basis of our data, we propose that acetylation-mediated regulation, whereby acetylation of p53 abrogates its association with the acidic-domain-containing cofactors, can be expanded to a general mode of post-translational control for protein interactions that involve other acidic-domain-containing factors and their ligands, which can be modified by acetylation.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 8 January; accepted 19 August 2016.
Published online 14 September 2016.

1. Zhao, S. *et al.* Regulation of cellular metabolism by protein lysine acetylation. *Science* **327,** 1000–1004 (2010).
2. Choudhary, C. *et al.* Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science* **325,** 834–840 (2009).
3. Kim, S. C. *et al.* Substrate and functional diversity of lysine acetylation revealed by a proteomics survey. *Mol. Cell* **23,** 607–618 (2006).
4. Gu, W. & Roeder, R. G. Activation of p53 sequence-specific DNA binding by acetylation of the p53 C-terminal domain. *Cell* **90,** 595–606 (1997).
5. Dhalluin, C. *et al.* Structure and ligand of a histone acetyltransferase bromodomain. *Nature* **399,** 491–496 (1999).
6. Marmorstein, R. & Zhou, M. M. Writers and readers of histone acetylation: structure, mechanism, and inhibition. Cold Spring Harb. *Perspect. Biol.* **6,** a018762 (2014).
7. Kim, K. *et al.* Vpr-binding protein antagonizes p53-mediated transcription via direct interaction with H3 tail. *Mol. Cell. Biol.* **32,** 783–796 (2012).
8. Zhao, L. Y. *et al.* Negative regulation of p53 functions by Daxx and the involvement of MDM2. *J. Biol. Chem.* **279,** 50566–50579 (2004).
9. Nair, B. C. *et al.* Proline, glutamic acid and leucine-rich protein-1 is essential for optimal p53-mediated DNA damage response. *Cell Death Differ.* **21,** 1409–1418 (2014).
10. Li, T. *et al.* Tumor suppression in the absence of p53-mediated cell-cycle arrest, apoptosis, and senescence. *Cell* **149,** 1269–1283 (2012).
11. Jiang, L. *et al.* Ferroptosis as a p53-mediated activity during tumour suppression. *Nature* **520,** 57–62 (2015).
12. Simeonova, I. *et al.* Mutant mice lacking the p53 C-terminal domain model telomere syndromes. *Cell Reports* **3,** 2046–2058 (2013).

13. Hamard, P. J. *et al.* The C terminus of p53 regulates gene expression by multiple mechanisms in a target- and tissue-specific manner *in vivo*. *Genes Dev*. **27,** 1868–1885 (2013).
14. von Lindern, M. *et al.* Can, a putative oncogene associated with myeloid leukemogenesis, may be activated by fusion of its 3′ half to different genes: characterization of the set gene. *Mol. Cell. Biol.* **12,** 3346–3355 (1992).
15. Kim, J. Y. *et al.* Inhibition of p53 acetylation by INHAT subunit SET/TAF-Iβ represses p53 activity. *Nucleic Acids Res*. **40,** 75–87 (2012).
16. Tang, Z. *et al.* SET1 and p300 act synergistically, through coupled histone modifications, in transcriptional activation by p53. *Cell* **154,** 297–310 (2013).
17. Jin, Q. *et al.* Distinct roles of GCN5/PCAF-mediated H3K9ac and CBP/p300-mediated H3K18/27ac in nuclear receptor transactivation. *EMBO J*. **30,** 249–262 (2011).
18. Kruse, J. P. & Gu, W. Modes of p53 regulation. *Cell* **137,** 609–622 (2009).
19. Vousden, K. H. & Prives, C. Blinded by the light: The growing complexity of p53. *Cell* **137,** 413–431 (2009).
20. Berger, S. L. Keeping p53 in check: a high-stakes balancing act. *Cell* **142,** 17–19 (2010).
21. Matsumoto, K., Nagata, K., Okuwaki, M. & Tsujimoto, M. Histone- and chromatin-binding activity of template activating factor-I. *FEBS Lett*. **463,** 285–288 (1999).
22. Kim, K. B. *et al.* Inhibition of Ku70 acetylation by INHAT subunit SET/TAF-Iβ regulates Ku70-mediated DNA damage response.*Cell. Mol. Life Sci.* **71,** 2731–2745 (2014).
23. Chae, Y. C. *et al.* Inhibition of FoxO1 acetylation by INHAT subunit SET/TAF-Iβ induces p21 transcription. *FEBS Lett.* **588,** 2867–2873 (2014).
24. Kouzarides, T. Chromatin modifications and their function. *Cell* **128,** 693–705 (2007).
25. Cohen, H. Y. *et al.* Acetylation of the C terminus of Ku70 by CBP and PCAF controls Bax-mediated apoptosis. *Mol. Cell* **13,** 627–638 (2004).
26. Daitoku, H. *et al.* Silent information regulator 2 potentiates Foxo1-mediated transcription through its deacetylase activity. *Proc. Natl Acad. Sci. USA* **101,** 10042–10047 (2004).
27. Krummel, K. A., Lee, C. J., Toledo, F. & Wahl, G. M. The C-terminal lysines fine-tune P53 stress responses in a mouse model but are not required for stability control or transactivation. *Proc. Natl Acad. Sci. USA* **102,** 10188–10193 (2005).
28. Feng, L., Lin, T., Uranishi, H., Gu, W. & Xu, Y. Functional analysis of the roles of posttranslational modifications at the p53 C terminus in regulating p53 stability and activity. *Mol. Cell. Biol.* **25,** 5389–5395 (2005).
29. UniProt Consortium. UniProt: a hub for protein information. *Nucleic Acids Res.* **43,** D204–D212 (2015).
30. Li, Y. *et al.* Accurate *in silico* identification of species-specific acetylation sites by integrating protein sequence-derived and functional features. *Sci. Rep.* **4,** 5765 (2014).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank F. Giancotti, X. Yang, R. K. Vadlamudi and W. An for providing reagents for this work. We also thank R. Baer for discussion and suggestions. This work was supported by the National Cancer Institute of the National Institutes of Health under Award 5R01CA193890, 5RO1CA190477, 5RO1CA085533 and 2P01CA080058 to W.G. and GM030518 and CA121852 to B.H. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Author Contributions The experiments were conceived and designed by D.W., N.K., G.L. and W.G. The experiments were performed mainly by D.W. and N.K. Bioinformatic analysis was performed by G.L. Mass spectrometry analysis was performed by W.L. The xenograft assay was performed by D.W. and L.J. Data were analysed and interpreted by D.W., N.K., G.L., W.-G.Z., J.Q., B.H. and W.G. The manuscript was written by D.W., N.K., G.L. and W.G.

Author Information RNA-seq data is available through NCBI Gene Expression Omnibus (GEO) database with the accession number GSE83635. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to W.G. (wg8@cumc.columbia.edu).

METHODS

General data reports. No statistical methods were used to pre-evaluate the sample size in this study. The experiments (including animal experiments) were not randomized. The investigators were not blinded to experiments. No samples/data were excluded except any obviously unhealthy xenografted mice.

Cell culture, plasmid generation, transfection and reagent treatment. H1299, U2OS, MCF7, H460 and HCT116 cell lines were cultured in DMEM supplemented with 10% (vol/vol) FBS. The SU-DHL-5 cell line was cultured in IMDM supplemented with 10% (vol/vol) FBS. MEFs were cultured in DMEM supplemented with 10% (vol/vol) heat-inactivated FBS. All the cell lines were obtained from ATCC and have been proven to be negative for mycoplasma contamination. No cell lines used in this work were listed in the ICLAC database. The cell lines were freshly thawed from the purchased seed cells and were cultured for no more than 2 months. The morphology of cell lines was checked every week and compared with the ATCC cell line image to avoid cross-contamination or misuse of cell lines. SET stable knockdown cells were generated by lentivirus-based infection of shRNA. *SET* cDNA was purchased from Addgene (Plasmid number 24998) and the full-length cDNA or the various fragments were sub-cloned into pWG-F-HA, pCMV-Myc or PGEX-2TL vectors. Each p53 plasmid was generated by sub-cloning human *p53* cDNA (including full-length or various fragments) into pWG-F-HA, pcDNA3.1 or PGEX-2TL vectors. The point-mutation constructs (including p53-KR and -KQ) were generated by using a site-directed mutagenesis Kit (Stratagene, 200521). Introduction of the expressing construct and siRNA transfection were performed by Lipofectamine 2000 (Invitrogen, 11668-019) according to the manufacturer's protocol. To transfer oligos into SU-DHL-5 cells, we used electroporation following the manufacturer's protocol (Lonza PBC3-00675). The DNA damage inducer doxorubicin was used at 1 μM for 24 h. The proteasome inhibitor epoxomicin was used at 100 nM for 6 h. Cells were treated with TSA (1 μM) and nicotinamide (5 mM) for 6 h to inhibit HDAC activity in the assays in which p53 acetylation needed to be maintained. Ad–GFP and Ad–Cre–GFP viruses were purchased from Vector Biolabs (Catalogue numbers 1761 and 1710).

Mouse model. To generate the knock-in mice, W4/129S6 mouse embryonic stem (ES) cells (Taconic) were electroporated with a targeting vector containing homologous regions flanking the mouse p53 exon 11, in which all 7 lysines were mutated to glutamines ($p53^{KQ}$ allele). A neomycin-resistance gene cassette flanked by two LoxP sites (LNL) was inserted into intron 10 to allow selection of targeted ES cell clones with G418. ES cell clones were screened by Southern blotting with EcoRI-digested genomic DNA, using a probe generated from PCR amplification in the region outside the homologous region in the targeting vector. The correctly targeted ES cell clones containing the K-to-Q mutations were injected into C57BL/6 blastocysts, which were then implanted into pseudopregnant females to generate chimaeras. Germ-line transmission was accomplished by breeding chimaeras with C57BL/6 mice. Subsequently, mice containing the targeted allele were bred with Rosa26-Cre mice to remove the LNL cassette and to generate mice with only the K-to-Q mutations. To confirm the mutations inserted in $p53^{+/KQ}$ mice, we sequenced p53 cDNA derived from mRNA isolated from $p53^{+/KQ}$ spleen. All seven K-to-Q mutations were confirmed and no additional mutations were found. The offspring were genotyped by PCR using the following primer set, forward: 5′-GGGAGGATAAACTGATTCTCAGA-3′, reverse: 5′-GATGGCTTCTACTATGGGTAGGGAT-3′.

To generate a *Set* conditional knockout mouse, exon 2 of the *Set* gene was floxed and deletion of exon 2 resulted in a frameshift and the truncation of the C-terminal domain. The targeting vector of Set contained 10 kb genomic DNA spanning exon 2; a neomycin-resistance gene cassette and loxP sites were inserted flanking exon 2. To increase targeting frequency, a diphtheria toxin A cassette was inserted at the 3′ end of the targeting vector to reduce random integration of the modified *Set* genomic DNA. A new BglII restriction site was also inserted to facilitate Southern blot screening. Of the 200 mouse ES cell clones screened, eight were identified to have integrated the floxed exon 2 by Southern blot using a 5′ probe, which detects a 14-kb band for the wild-type allele and an 11-kb band for the floxed exon 2 allele (Set^{flox}). Two of the clones were then injected into blastocysts to generate *Set* chimaera mice and they were bred to produce germ-line transmission of the floxed exon 2 allele. $Set^{flox/+}$ mice were intercrossed to generate Set homozygous conditional knockout mice ($Set^{flox/flox}$).

Maintenance and experimental procedures of mice were approved by the Institutional Animal Care and Use Committee (IACUC) of Columbia University.

***In vitro* binding assay.** For the *in vitro* peptide binding assay: equal amounts of each synthesized biotin-conjugated peptide (made as column or as batch) were incubated with highly concentrated HeLa nuclear extract (NE) or purified proteins for 1 h or overnight at 4 °C. After washing with BC100 buffer (20 mM Tris-HCl pH 7.9, 100 mM NaCl, 10% glycerol, 0.2 mM EDTA, 0.1% triton X-100) three times, the binding components were eluted in high-salt buffer (20 mM Tris-HCl pH 7.9, 1,000 mM NaCl, 1% DOC, 10% glycerol, 0.2 mM EDTA, 0.1% triton X-100) or by boiling with 1 × Laemmli buffer for further analysis. For the *in vitro* GST-fusion protein binding assay: *Escherichia coli* containing GST or GST-fusion protein expressing constructs were grown in a shaking incubator at 37 °C until the OD_{600} was about 0.6. Next 0.1 mM IPTG was added and the *E. coli* were incubated at 25 °C for 4 h or overnight, to induce GST or GST-fusion protein expression. After purification by GST·Bind Resin (Novagen, 70541), equal amounts of immobilized GST or GST-fusion proteins were incubated with other purified proteins for 1 h at 4 °C, followed by washing with BC100 buffer three times. The binding components were eluted by boiling with 1 × Laemmli buffer and were analysed by western blot.

Co-immunoprecipitation assay (Co-IP). Whole cellular extracts (WCE) were prepared in BC100 buffer with sonication. Nuclear extract (NE) was prepared by sequentially lysing cells with HB buffer (20 mM Tris-HCl pH 7.9, 10 mM KCl, 1.5 mM $MgCl_2$, 1 mM PMSF, 1 × protease inhibitor (Sigma)) for the cytosolic fraction and BC400 buffer (20 mM Tris-HCl pH 7.9, 400 mM NaCl, 10% Glycerol, 0.2 mM EDTA, 0.5% triton X-100, 1 mM PMSF, 1 × protease inhibitor) for nuclear fraction. The salt concentration of NE was adjusted to 100 mM. 2 μg of the indicated antibody (or 20 μl Flag M2 Affinity Gel (Sigma, A2220)) was added into WCE or NE and incubated overnight at 4 °C, followed by addition of 20 μl protein A/G agarose (Santa Cruz, sc-2003; only for IP with unconjugated antibodies mentioned above) for 2 h. After washing with BC100 buffer three times, the binding components were eluted using Flag peptide (Sigma, F3290), 0.1% trifluoroacetic acid (TFA, Sigma, 302031) or by boiling with 1 × Laemmli buffer, and were analysed by western blot.

Purification of Ub-, Sumo- or Nedd-p53 conjugates from cells. For preparation of Ub-p53: H1299 cells were co-transfected with p53, MDM2 and 6 × HA-Ub (human) expressing plasmids for 48 h. The cells were lysed with Flag lysis buffer (50 mM Tris-HCl pH 7.9, 137 mM NaCl, 10 mM NaF, 1 mM Na_3VO_4, 10% glycerol, 0.5 mM EDTA, 1% triton X-100, 0.2% sarkosyl (sodium lauroyl sarcosinate), 0.5 mM DTT, 1 mM PMSF, 1 × protease inhibitor) and total Ub-conjugated proteins were purified by anti-HA-agarose (Sigma, A2095) and eluted by 1 × HA peptide (Sigma I2149). For the preparation of Sumo-p53 or Nedd-p53: H1299 cells were co-transfected with p53, MDM2 (only for Nedd-p53 preparation) and 6 × His-HA-Sumo1 (human) or 6 × His-HA-Nedd8 (human) expressing plasmids for 48 h. The cells were lysed with guanidine lysis buffer (6 M guanidin-HCl, 0.1 M Na_2HPO_4, 6.8 mM NaH_2PO_4, 10 mM Tris-HCl pH 8.0, 0.2% triton-X100, freshly supplemented with 10 mM β-mercaptoethanol and 5 mM imidazole) with mild sonication. After overnight pull-down by Ni^+-NTA agarose (Qiagen 30230), the binding fractions were sequentially washed with guanidine lysis buffer, urea buffer I (8 M urea, 0.1 M Na_2HPO_4, 6.8 mM NaH_2PO_4, 10 mM Tris-HCl pH 8.0, 0.2% triton-X100, freshly supplemented with 10 mM β-mercaptoethanol and 5 mM imidazole) and urea buffer II (8 M urea, 18 mM Na_2HPO_4, 80 mM NaH_2PO_4, 10 mM Tris-HCl pH 6.3, 0.2% triton-X100, freshly supplemented with 10 mM β-mercaptoethanol and 5 mM imidazole). Precipitates were eluted in elution buffer (0.5 M imidazole, 0.125 M DTT). All purified proteins were dialysed against BC100 buffer before use in the subsequent pull-down assay. After the pull-down assay, the interaction between SET and each p53-conjugate was detected by western blot with anti-p53 (DO-1) antibody.

Mass spectrometry assay. The protein complex was separated by SDS–PAGE and stained with GelCode Blue reagent (Pierce, 24592). The visible band was cut and digested with trypsin and then subjected to liquid chromatography (LC)-MS/MS analysis.

Luciferase assay. A firefly reporter (p21-Luci reporter) and a Renilla control reporter were co-transfected with indicated constructs in H1299 cells for 48 h and the relative luciferase activity was measured by dual-luciferase assay protocol (Promega, E1910).

Electrophoretic mobility shift assay. Highly purified p53 or SET was incubated with a ^{32}P-labelled probe (160 bp) containing the p53-binding element of the *p21* promoter in 1× binding buffer (10 mM HEPES, pH 7.6, 40 mM NaCl, 50 μM EDTA, 6.25% glycerol, 1 mM $MgCl_2$, 1 mM spermidine, 1 mM DTT, 50 ng μl^{-1} BSA, 5 ng μl^{-1} sheared single strand salmon DNA) for 20 min at room temperature (RT). For the super-shift assay, α-p53 or α-SET antibody was pre-incubated with purified p53 and SET in the reaction system without probe for 30 min at RT and then the probe was added for a further 20 min. The complex was analysed by 4% Tris-Borate-EDTA buffer–polyacrylamide gel electrophoresis (TBE–PAGE) and visualized by autoradiography. The probe was obtained by PCR, labelled by T4 kinase (NEB, M0201S) and purified by Bio-Spin column (Bio-Rad, 732-6223).

Chromatin immunoprecipitation (ChIP) assay. Cells were fixed with 1% formaldehyde for 10 min at room temperature and lysed with ChIP lysis buffer (50 mM Tris-HCl pH 8.0, 5 mM EDTA, 1% SDS, 1× protease inhibitor) for 10 min at 4 °C.

After sonication, the lysates were centrifuged, and the supernatants were collected and pre-cleaned by salmon sperm DNA saturated protein A agarose (Millipore, 16-157) in dilution buffer (20 mM Tris-HCl pH 8.0, 2 mM EDTA, 150 mM NaCl, 1% triton X-100, 1× protease inhibitor) for 1 h at 4 °C. The pre-cleaned lysates were aliquoted equally and incubated with indicated antibodies overnight at 4 °C. Saturated protein A agarose was added into each sample and incubated for 2 h at 4 °C. The agarose was washed with TSE I (20 mM Tris-HCl pH 8.0, 2 mM EDTA, 150 mM NaCl, 0.1% SDS, 1% triton X-100), TSE II (20 mM Tris-HCl pH 8.0, 2 mM EDTA, 500 mM NaCl, 0.1% SDS, 1% triton X-100), buffer III (10 mM Tris-HCl pH 8.0, 1 mM EDTA, 0.25 M LiCl, 1% DOC, 1% NP40), and buffer TE (10 mM Tris-HCl pH 8.0, 1 mM EDTA), sequentially. The binding components were eluted in 1% SDS and 0.1 M $NaHCO_3$ and reverse cross-linkage was performed at 65 °C for at least 6 h. DNA was extracted using the PCR purification Kit (Qiagen, 28106). Real-time PCR was performed to detect relative enrichment of each protein or modification on indicated genes.

Cell growth assay. Approximately 10^5 MEFs or U2OS cells, as indicated in each figure, were seeded into 6-well plates with three replicates. Their cell growth was monitored on consecutive days, as indicated, by using the Countess automated cell counter (Invitrogen) or by staining with 0.1% crystal violet. For quantitative analysis of the crystal violet staining, the crystal violet was extracted from cells using 10% acetic acid and the relative cell number was measured by detecting the absorbance at 590 nm.

Xenograft model. 10^6 HCT116-derived cells, as indicated in each figure, were mixed with Matrigel (Corning, 354248) in a 1:1 ratio in a total volume of 200 μl. The cell–matrix complex was subcutaneously injected into nude mice (NU/NU; 8 weeks old; female; strain 088; Charles River). After 3 weeks, the mice were killed and weight of the tumours was measured. The experimental procedures were approved by the Institutional Animal Care and Use Committee (IACUC) of Columbia University. None of the experiments were exceeded the limit for tumour burden (10% of total bodyweight or 2 cm in diameter).

RT–qPCR. Total RNA was extracted by TRIzol (Invitrogen, 15596-026) and precipitated in ethanol. 1 μg of total RNA was reverse transcribed into cDNA using the SuperScript III First-Strand Synthesis SuperMix (Invitrogen, 11752-50). The relative expression of each target was measured by qPCR and the data were normalized by the relative expression of *GAPDH* or *ActB*.

Immunohistochemistry (IHC). FFPE sections of mouse brain tissue samples were stained with indicated antibodies and visualized by DAB exposure.

Protein purification. The Flag-tagged p53 or SET construct was transfected into H1299 cells for 48 h and the cells were lysed in Flag lysis buffer. After centrifugation, the Flag M2 Affinity Gel was added to supernatant and incubated for 1h at 4 °C. After washing with Flag lysis buffer six times, the purified proteins were eluted with Flag peptide. For purification of acetylated p53, the construct CBP was co-transfected with the p53 vector for 48 h. TSA and nicotinamide were added into the medium for the last 6 h and the cells were harvested in Flag lysis buffer supplemented with TSA and nicotinamide. The C-terminal unacetylated p53 was removed by p53-PAb421 antibody and then the acetylated p53 was purified as described above.

***In vitro* acetylation assay.** 0.5 μg recombinant H3 was incubated with 20 ng purified p300 in 1× HAT buffer (50 mM Tris-HCl, pH 7.9; 1 mM DTT; 10 mM sodium butyrate, 10% glycerol) containing 0.1 mM Ac-CoA for 30 min at 30 °C. After the reaction, the products were assayed by western blot with indicated antibodies. To measure the effect of SET on p300-mediated H3 acetylation, H3 and purified SET (1 μg) were pre-incubated in 1× HAT buffer for 20 min at room temperature before addition of the other components (p300 and Ac-CoA) for the subsequent *in vitro* acetylation assay.

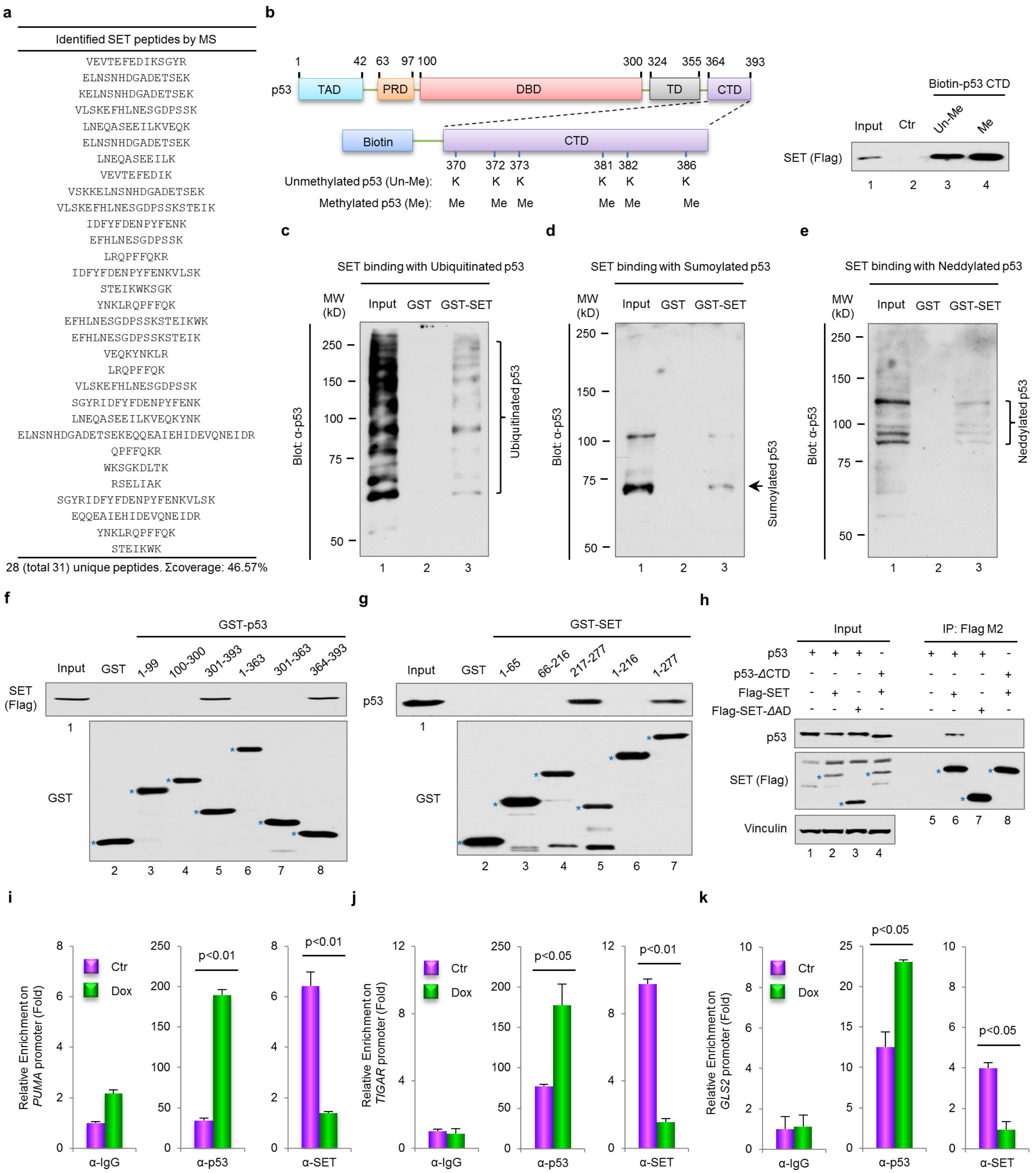
Generation of the p53 knockout (p53-KO) cell line using the CRISPR/Cas9 technique. Cells were transfected with constructs expressing Cas9-D10A (Nickase) and control sgRNAs or sgRNAs targeting p53 exon3 (Santa Cruz: sc-437281 for control; sc-416469-NIC for targeting of p53). After 48 h of transfection, cells were suspended, diluted and re-seeded to ensure single clone formation. More than 30 clones were picked up and the expression of p53 in each single clone was evaluated by western blot with both α-p53 (DO-1) and α-p53 (FL-393) antibodies. Further verification of positive clones was done by sequencing the genomic DNA to make sure that the functional genomic editing occured (insertion or deletion-mediated frame-shift of the *p53* open reading frame (ORF)). Two (U2OS) or three (HCT116) clones were finally selected for subsequent experiments. The p53 knockout-mediated effect was verified to be reproducible in these independent clones. The targeting sequences of p53 loci for the sgRNAs were: 1) TTGCCGTCCCAAGCAATGGA; 2) CCCCGGACGATATTGAACAA.

RNA-seq. U2OS (CRISPR Ctr or CRISPR p53-KO) cells were transfected with control siRNA or SET-specific siRNA (three oligos) for 4 days. Each sample group had at least two biological replicates. Total RNA was prepared using TRIzol (Invitrogen, 15596-026). The RNA quality was evaluated by Bioanalyzer (Agilent) and confirmed that the RIN > 8. Before performing RNA-seq analysis, a small aliquot of each sample was analysed by RT–qPCR to confirm SET knockdown efficiency. RNA-seq analysis was performed at the Columbia Genome Center. Specifically, from total RNA samples, mRNAs were enriched by poly-A pull-down and then processed for library preparation by using the Illumina TruSeq RNA prep kit (Illumina RS-122-2001). Libraries were then sequenced using the Illumina HiSeq2000. Samples were multiplexed in each lane and yielded targeted number of single-end 100-bp reads for each sample. RTA (Illumina) was used for base calling and bcl2fastq (version 1.8.4) was used for converting BCL to fastq format, coupled with adaptor trimming. Reads were mapped to a reference genome (Human: NCBI/build37.2) using TopHat (version 2.0.4). Relative abundance of genes and splice isoforms were determined using Cufflinks (version 2.0.2) using the default settings. Differentially expressed genes were tested under various conditions using DEseq, an R package based on a negative binomial distribution that models the number reads from RNA-seq experiments and tests for differential expression. To further analyse the differentially expressed genes in a more reliable interval, the following filter strategies were applied: 1) the average of FPKM (Fragments per kilobase of transcript per million mapped reads) in either sample group exceeded 0.1; 2) the fold change between the CRISPR Ctr/si-Ctr group and the CRISPR Ctr/si-SET group exceeded 2; 3) the *P* value between the CRISPR Ctr/si-Ctr group and the CRISPR Ctr/si-SET group < 0.01.

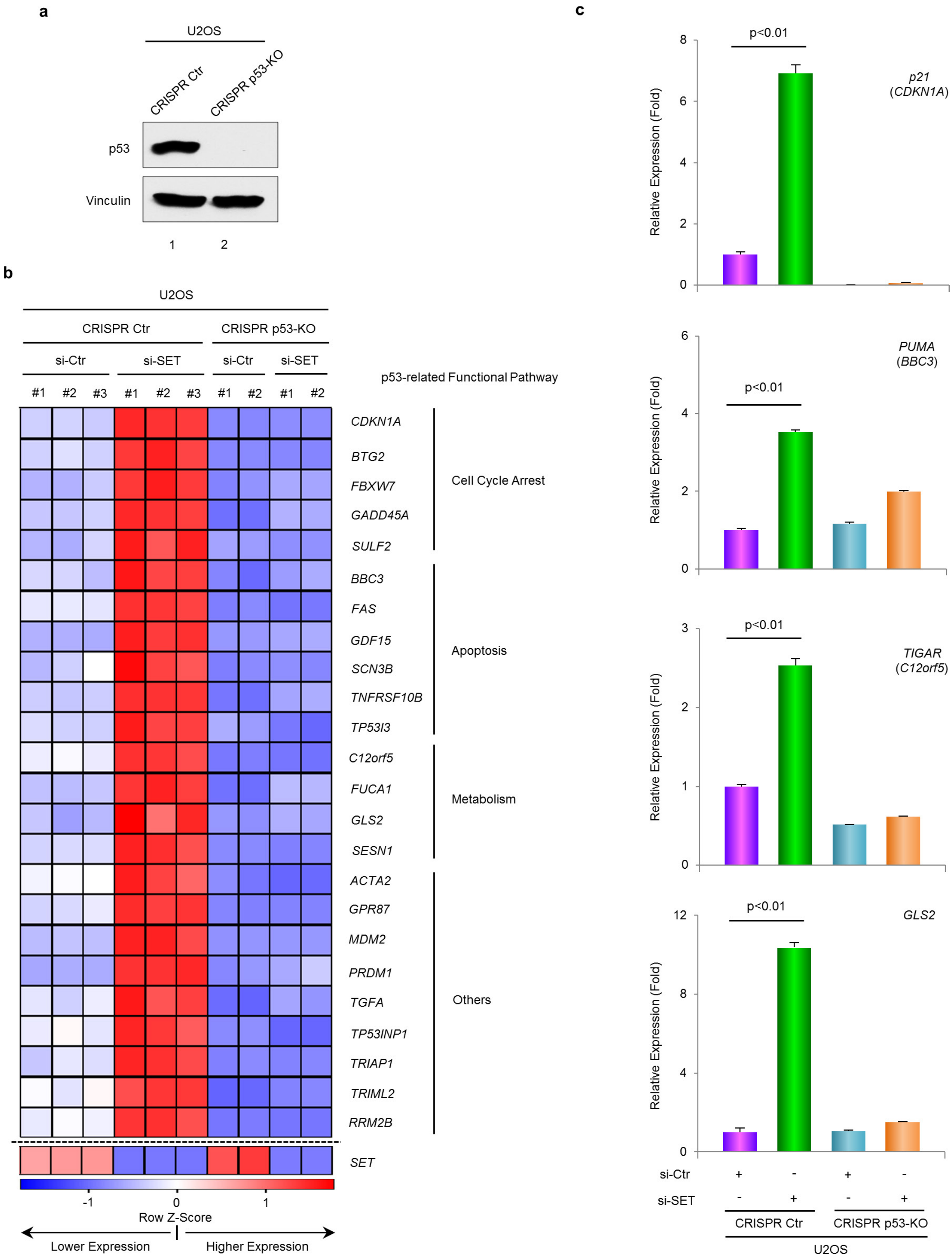
To retrieve potential p53 target genes which were repressed by SET in a p53-dependent manner, we searched the filtered RNA-seq results using the following strategies: 1) the expression level in the CRISPR Ctr/si-SET group was at least 2-fold higher than that in the CRISPR Ctr/si-Ctr group; 2) the expression level in the CRISPR Ctr/si-SET group was at least 2-fold higher than that in the CRISPR p53-KO/si-SET group. The filtered genes which were also verified as p53 target genes from the literature were collected and presented as a heatmap.

Bioinformatic analysis. For the discovery of acidic domains in the human proteome: our motif-finding algorithm initially searched for sequence motifs with a minimum acidic composition of 76% using a sliding window of 36 residues, as dictated by experimental results. Motifs found to be partially overlapping were merged into single motifs. Flanking non-acidic residues were subsequently cropped-out from the final motif. Motif discovery was carried out using the UniProt database, which contains 20,187 canonical human proteins, that have been manually annotated and reviewed. For prediction of proteins that bound acidic domain-containing proteins and were regulated by acetylation: we identified proteins that can potentially bind long acidic domains in a similar way to p53: using a K-rich region whose binding properties can be regulated by acetylation. We used the training set assembled in SSPKA, which combines lysine acetylation annotations from multiple resources obtained either experimentally or in the scientific literature. This dataset individually lists all annotated acetylation sites for a given protein. We generated acetylation motifs with multiple acetylation sites by clustering those sites found to within a maximum distance of 11 residues in sequence. Following this, we searched for acetylation motifs with five or more lysines where at least three of them are annotated as acetylation sites.

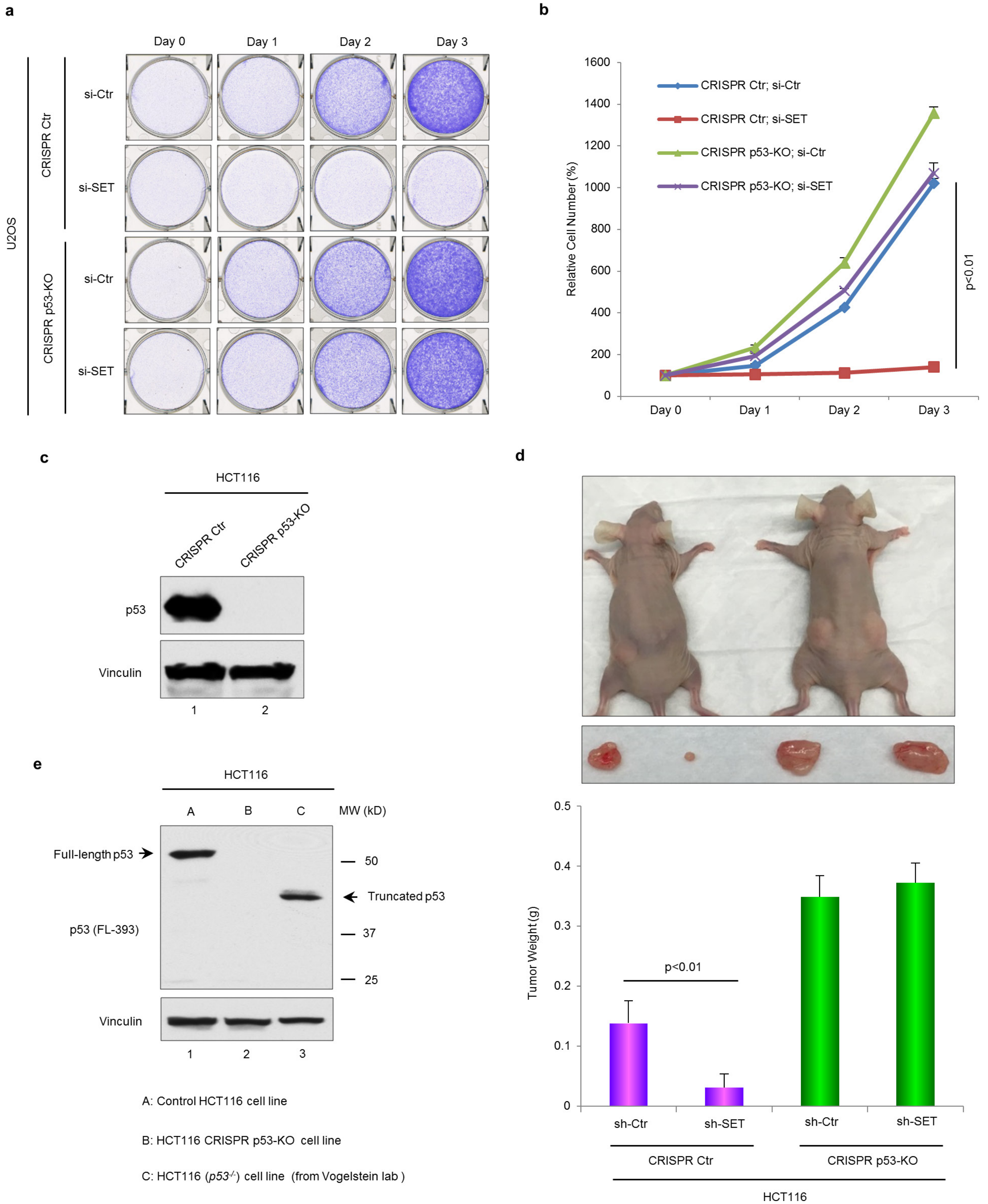
Statistical analysis. Results are shown as means ± s.d. Statistical significance was determined by using a two-tailed, unpaired Student *t*-test in all figures except those described below. In Fig. 1g, significance was determined by one-way ANOVA with a Bonferroni post hoc test. In Fig. 2d and g and Extended Data Figs 2c, 3b, d, 4f and 7h, statistical significance was measured by two-way ANOVA with a Bonferroni post hoc test. All statistical analysis was performed using GraphPad Prism software. $P < 0.05$ was denoted as statistically significant.



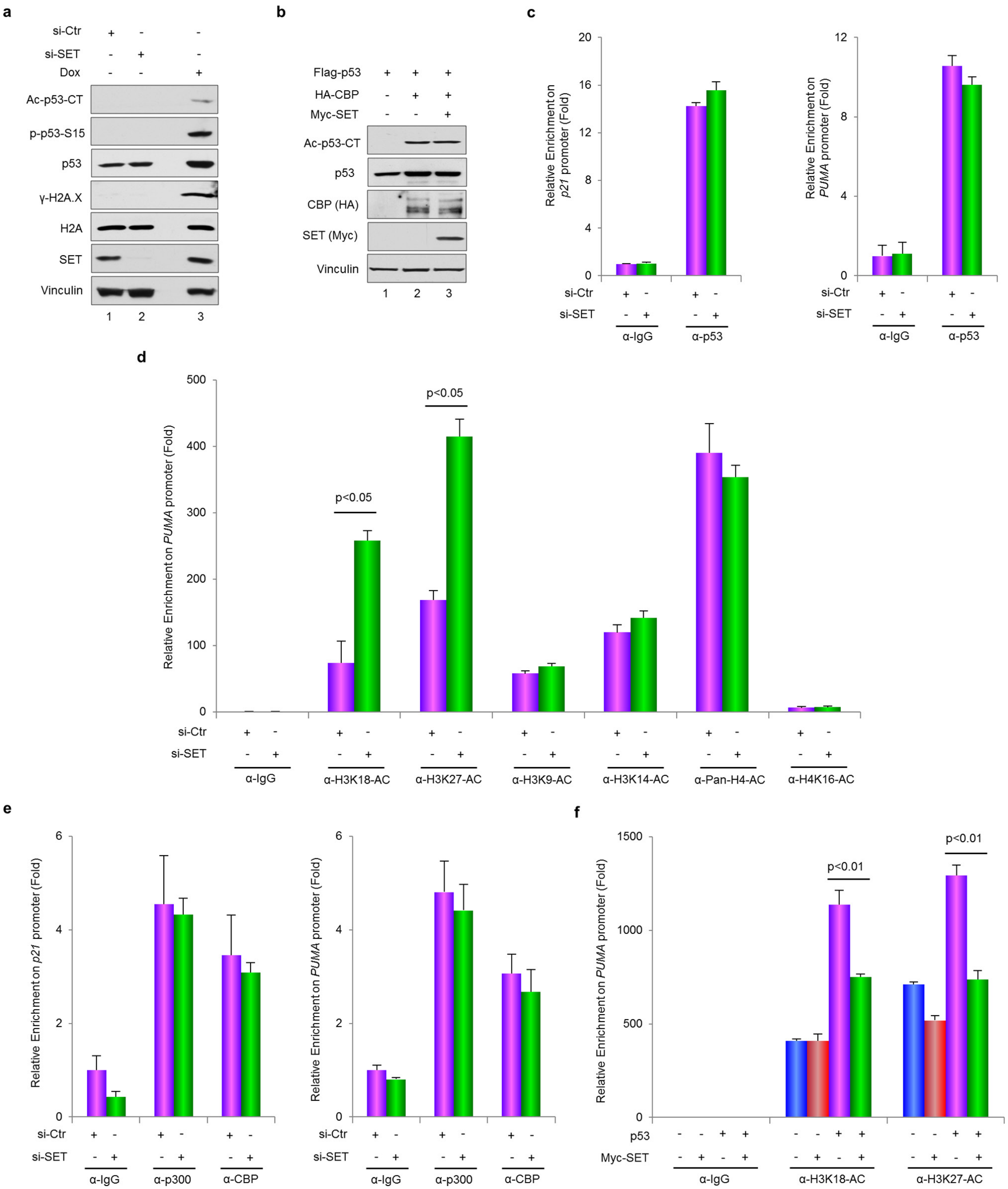
Extended Data Figure 1 | Further analysis of p53–SET interaction. **a**, A list of SET peptides identified by mass spectrometry. **b**, *In vitro* binding assay of methylated p53 CTD and purified SET. **c**–**e**, *In vitro* binding assay between SET and the purified ubiquitinated, sumoylated or neddylated forms of p53. **f**, **g**, Western blot analysis of p53 and SET domains for their interaction. *In vitro* binding assay was performed by incubating immobilized GST, GST–p53 or GST–SET with each purified SET or p53 protein, as indicated. **h**, Western blot analysis of the interaction between p53 and SET in cells. H1299 cells were co-transfected with indicated constructs and the nuclear extract was analysed by co-IP assay. **i**–**k**, ChIP analysis of p53 or SET recruitment onto the *PUMA* (**i**), *TIGAR* (**j**) or *GLS2* (**k**) promoter. HCT116 cells were treated with or without 1 μM doxorubicin for 24 h and then the cellular extracts were analysed by ChIP assay with indicated antibodies. Asterisks indicate the specific bands of indicated proteins. Error bars indicate mean ± s.d., $n = 3$ for technical replicates. Data are shown as representative of three experiments. Uncropped blots can be found in Supplementary Fig. 1.



Extended Data Figure 2 | RNA-seq analysis to identify genes regulated by p53–SET interplay. a, Western blot analysis of the expression of p53 in U2OS-derived CRISPR control cells or CRISPR p53-KO cells. **b**, Heat map of genes regulated by the p53–SET interplay. U2OS (CRISPR Ctr or CRISPR p53-KO) cells were transfected with control siRNA or SET-specific siRNA for 4 days and the total RNA was prepared for RNA-seq analysis with two or three biological replicates, as indicated. Known p53 target genes which were also repressed by SET in a p53-dependent manner were selected and presented as a heat map. The relative *SET* expression is shown in the last row of the heat map. **c**, qPCR validation of the genes regulated by the p53–SET interplay. Error bars indicate mean ± s.d., $n = 3$ for technical replicates. Data are shown as representative of three experiments. Uncropped blots can be found in Supplementary Fig. 1.



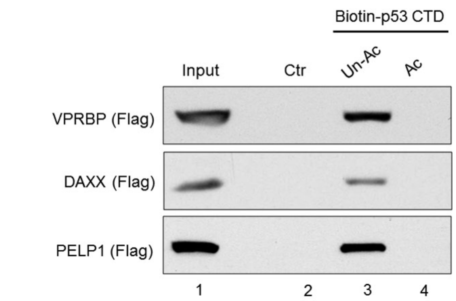
Extended Data Figure 3 | SET-mediated effects on cell proliferation and tumour growth. a, **b**, Representative image (**a**) or quantitative analysis (**b**) of the SET knockdown-mediated effect on cell growth of U2OS-derived CRISPR control cells or CRISPR p53-KO cells. **c**, Western blot analysis of the expression of p53 in HCT116-derived CRISPR control cells or CRISPR p53-KO cells. **d**, Xenograft analysis of the SET-mediated effect on tumour growth by HCT116-derived CRISPR control cells or CRISPR p53-KO cells. **e**, Western blot analysis of p53 expression in control or derived HCT116 cell lines, as indicated. Error bars indicate mean ± s.d., $n = 3$ in **b** or $n = 5$ in **d** for biological replicates. Uncropped blots can be found in Supplementary Fig. 1.

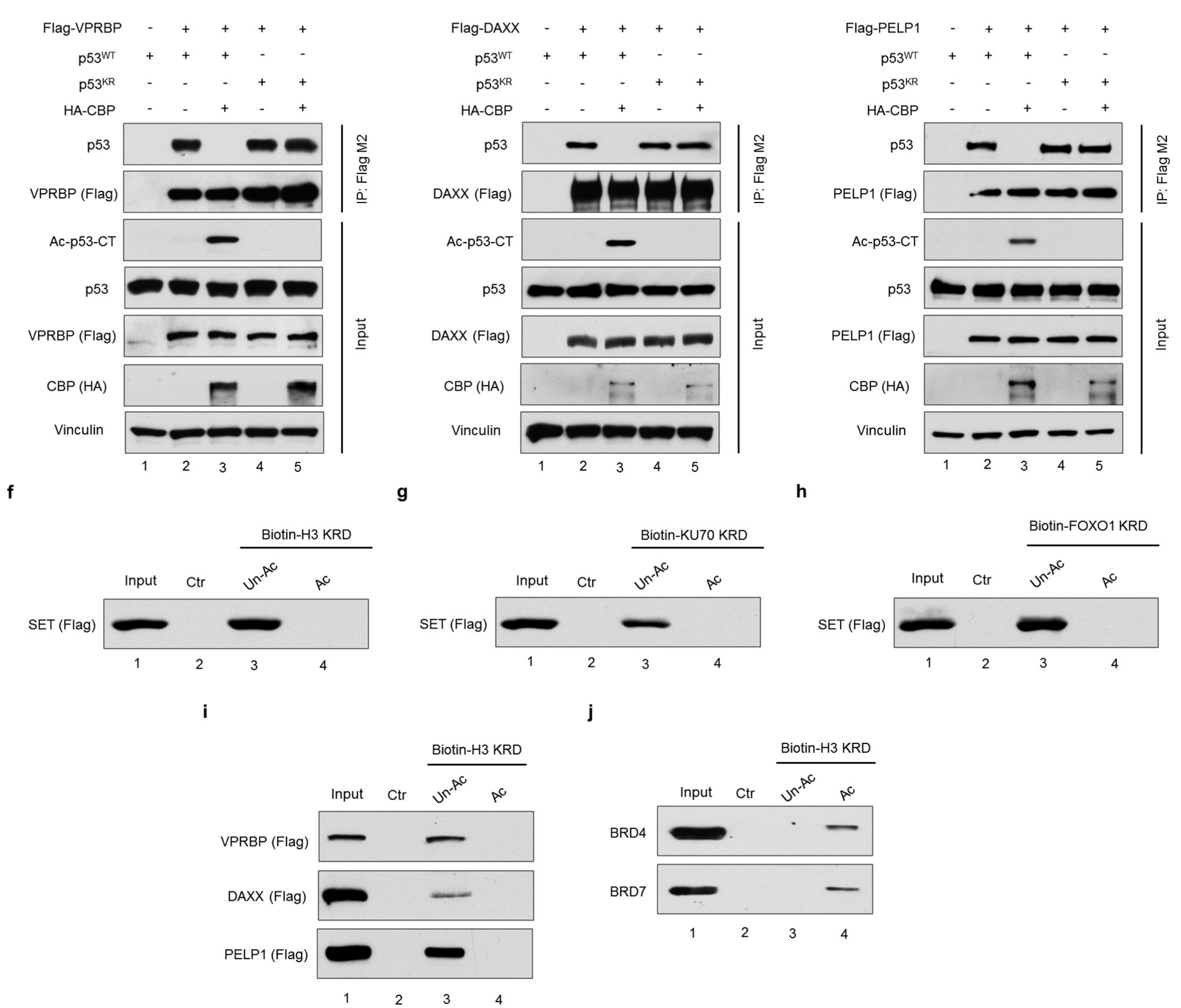


Extended Data Figure 4 | SET regulates histone modifications on p53 target promoter. a, Western blot analysis of the SET knockdown-mediated effect on the p53 C-terminal acetylation in HCT116 cells. Doxorubicin (Dox)-treated cells were also analysed in parallel as a positive control. **b**, Western blot analysis of the SET-mediated effect on the CBP-induced p53 C-terminal acetylation in H1299 cells. **c**, **e**, ChIP analysis of promoter-recruitment of p53 (**c**) or p300/CBP (**e**) upon SET depletion in HCT116 cells. **d**, ChIP analysis of the SET-knockdown-mediated effect on histone modifications in the *PUMA* promoter in HCT116 cells. **f**, ChIP analysis of the SET-mediated effect on p53-dependent H3K18 and H3K27 acetylation in the *PUMA* promoter. Error bars indicate mean ± s.d., $n = 3$ for technical replicates. Data are shown as representative of three experiments. Uncropped blots can be found in Supplementary Fig. 1.

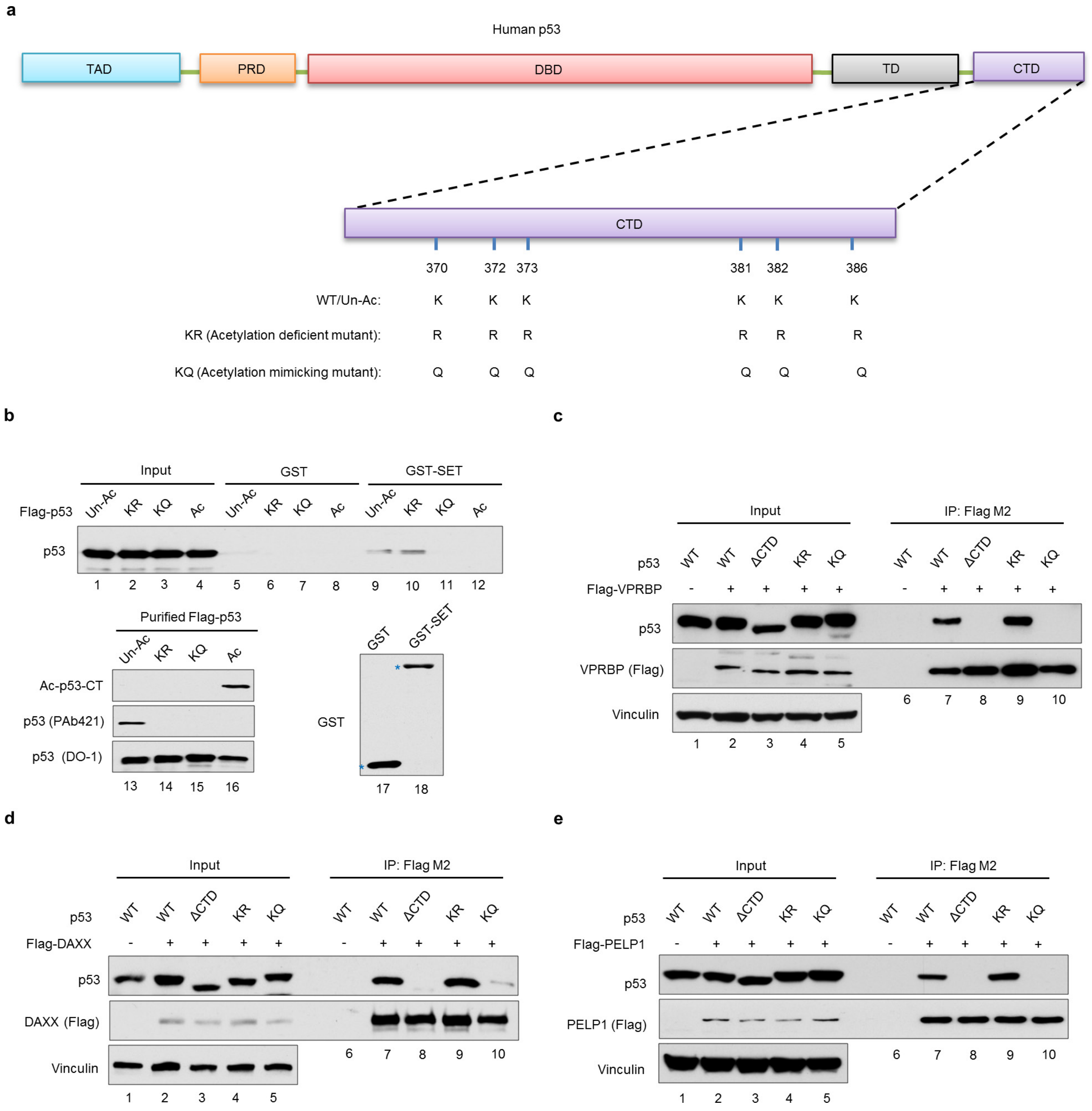
a

Protein Name	Acidic Domain Sequence	Length of Acidic Domain	Distribution Pattern of Acidic Amino Acid in Acidic Domain	Percentage of Acidic Amino Acid in Acidic Domain
SET	DDEEGEAEDDDDDDEEEEGLEDIDEEGDEDEGEEDDDE DEGEEGEEDEGEDD	52 (226-277)	Clustered	81%
VPRBP	EDEDEEEDQEEEEQEEEDDDEDDDDTDDLDELDTDQLL EAELEEDD	46 (1396-1441)	Clustered	76.10%
DAXX	ETDDEDDEESDEEEEEEEEEEEEEEATDSEEEEDLEQMQ EGQEDDEEEDEEEE	52 (434-485)	Clustered	78.80%
PELP1	DEEEEEEEEEEEEEEEEEEEEEEDFEEEEEDEEEYFEEE EEEEEEFEEEFEEEEGELEEEEEEEDEEEEEEELEEVED	76 (887-962)	Clustered	88.20%

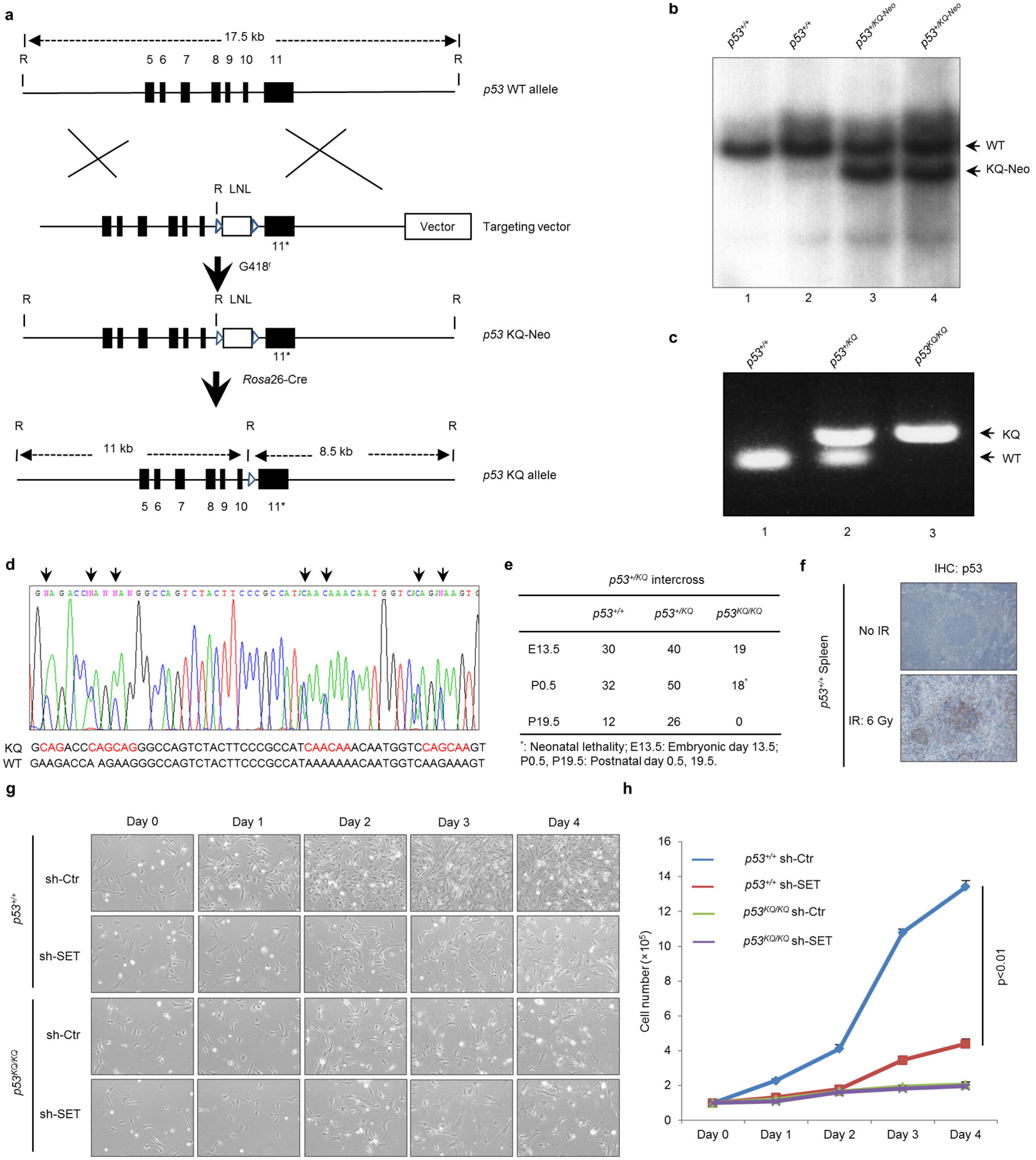




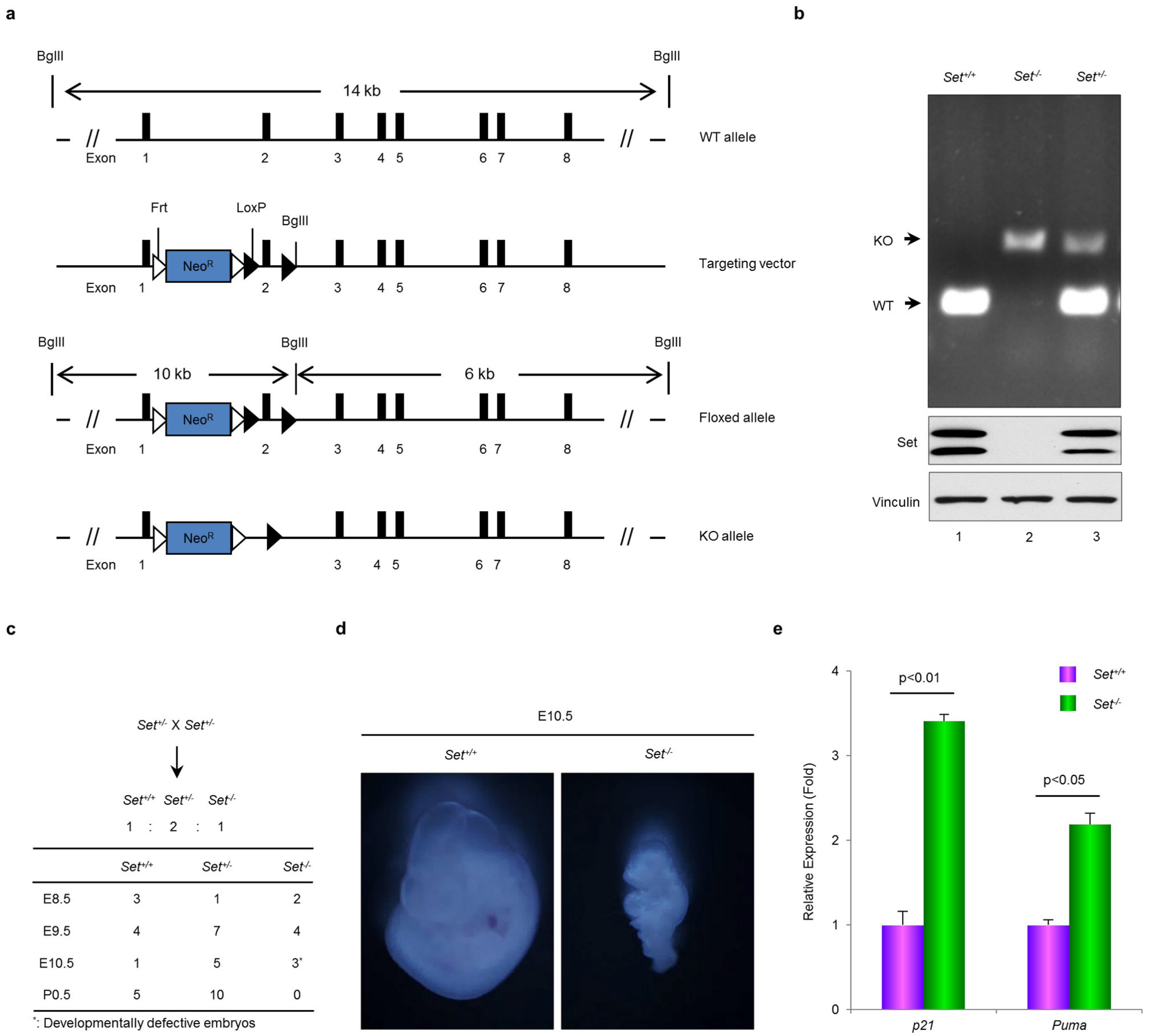
Extended Data Figure 5 | Acetylation regulates the interaction between acidic-domain-containing proteins and their acetylatable ligands. **a**, A summary table of characteristic features of the acidic-domain-containing proteins SET, VPRBP, DAXX and PELP1. The acidic amino acids are underlined. **b**, *In vitro* binding assay of p53 CTD and purified full-length VPRBP, DAXX or PELP1. **c**–**e**, Western blot analysis of the interaction between p53 and VPRBP (**c**), DAXX (**d**) or PELP1 (**e**) in the nuclear fraction of H1299 cells. **f**–**h**, *In vitro* binding assay between purified SET and KRD of H3 (**f**), KU70 (**g**) or FOXO1 (**h**). **i**, *In vitro* binding assay of the H3 KRD and purified VPRBP, DAXX or PELP1. **j**, *In vitro* binding assay of the H3 KRD and BRD4 or BRD7 (nuclear extract). Uncropped blots can be found in Supplementary Fig. 1.



Extended Data Figure 6 | p53^{KQ} mutant mimics acetylated p53. **a**, Schematic diagram of human unacetylated p53 and the acetylation-deficient and acetylation-mimicking mutants of p53. **b**, *In vitro* binding assay of SET and different types of p53, as indicated. **c**–**e**, Western blot analysis of the interaction between acidic-domain-containing proteins (**c**, VPRBP; **d**, DAXX; **e**, PELP1) and different types of p53 in cells. H1299 cells were co-transfected with indicated constructs, and the nuclear extract was analysed by Co-IP assay. Asterisks indicate the purified proteins. Uncropped blots can be found in Supplementary Fig. 1.



Extended Data Figure 7 | Generation of $p53^{KQ/KQ}$ mice. a, Schematic diagram of the gene targeting strategy to replace the p53 C-terminal 7 lysines with 7 glutamines in mouse *p53*. **b**, Southern blot screening of ES cells to identify $p53^{+/KQ}$ clones. **c**, PCR genotyping analysis of wild-type (110 bp), $p53^{+/KQ}$ heterozygous (110 bp and 150 bp), and $p53^{KQ/KQ}$ homozygous mice (150 bp only). **d**, Sequencing analysis of the transcripts prepared from the $p53^{+/KQ}$ heterozygous mouse spleen. **e**, A summary table of observed numbers of mice from $p53^{+/KQ}$ heterozygous intercrosses. **f**, Positive control for p53 staining in the IHC assay. The spleen tissue sections of $p53^{+/+}$ mice treated with or without 6 Gy γ-radiation was stained with p53 (CM-5) antibody. **g**, **h**, Representative image (**g**) or quantitative analysis (**h**) of SET-knockdown-mediated cell growth of $p53^{+/+}$ or $p53^{KQ/KQ}$ MEFs (P2). Error bars indicate mean ± s.d., $n = 3$ for biological replicates. Uncropped blots can be found in Supplementary Fig. 1.



Extended Data Figure 8 | Characterization of *Set* conditional knockout mice. a, Schematic diagram of the strategy to generate *Set* conditional knockout mice. **b**, Validation of *Set* knockout in embryos (E8.5) by genotyping and western blot analysis. **c**, A summary table of observed numbers of embryos or pups from $Set^{+/-}$ intercrosses. **d**, Representative pictures of $Set^{+/+}$ and $Set^{-/-}$ embryos (E10.5). **e**, qPCR analysis of the expression of p53 target genes in $Set^{+/+}$ and $Set^{-/-}$ embryos (E10.5). Error bars indicate mean $\pm$ s.d., $n = 3$ for technical replicates. Data are shown as representative of three experiments. Uncropped blots can be found in Supplementary Fig. 1.

Extended Data Table 1 | A list of human proteins containing acidic domains with a minimum percentage of acidic residues of 76% within a 36-residue window

	UniProt ID	Protein Name	Acidic Domain Position	Acidic Domain Sequence	Biological Function (GO)
Proteins Involved in Gene Expression Control through DNA Binding, Transcription Regulation and Chromatin Remodeling	Q8IZL8	Proline-, glutamic acid- and leucine-rich protein 1	886 - 963	DEEEEEEEEEEEEEEEEEEEEEEDFEEEEEDEEEYFEEEEEEEEFEEEFEEEEGELEEEEEEEDEEEEEEELEEVEDLE	Chromatin binding, Transcription factor binding, poly(A) RNA binding
	Q92688	Acidic leucine-rich nuclear phosphoprotein 32 family member B	156 - 232	DSDAEVDGVDEEEEDEEGEDEEDEDDEDGEEEEFDEEDDEDEDVEGDEDDDEVSEEEEEFGLDEEDEDEDEDEEEEE	Protein binding, Histone binding, RNA polymerase binding
	Q9UL68	Myelin transcription factor 1-like protein	107 - 169	EEYSEDNDEPGDEDEEDEEGDREEEEEIEEEDEDDDEDGEDVEDEEEEEEEEEEEEEEENED	Sequence-specific DNA binding, Transcription factor, Zinc binding
	Q01538	Myelin transcription factor 1	257 - 315	EEEDEEEEEEEEEEEEEDEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEAAPDVIFQED	Sequence-specific transcription factor, Zinc binding
	A1YPR0	Zinc finger and BTB domain-containing protein 7C	124 - 178	EIMEPGGDGGEEDDKEDDDDDEDDDDEEDEEEEEEEEDDDDDTEDFADQENLPD	Nucleic acid binding, Metal ion binding
	Q86V15	Zinc finger protein castor	1671 - 1724	ESSQEDEEEELELPEEEAEDDEDEDDDEDDDDEDDDEDDDDEDLRTDSEESLPE	DNA binding, Metal ion binding
	Q01105	Protein SET	236 - 289	DMDDEEGEGEEDDDDDEEEEGLEDIDEEGDEDEGEEDEDDDEGEEGEEDEGEDD	DNA binding, Histone binding, Phosphatase inhibitor
	P0DME0	Protein SETSIP	248 - 301	DDEEGGEDDDDDDDGDEGEEELEDIDEGDEDEGEEDEDDDEGEEGEEDEGEDD	Chromatin binding
	Q7Z6M4	Transcription termination factor 4, mitochondrial	332 - 380	DDKRASLDEDEDDDDEEDNDEDDNDEDDDDEDDDEAEDNDEDEDDDEEE	DNA binding, RNA binding, Protein binding
	Q6PL18	ATPase family AAA domain-containing protein 2	242 - 288	ESSEEGEDQEHEDDGEDEDDEDDDDDDDDDDDDDDEDDEDEEDGEEE	Histone binding, Chromatin binding, Hydrolase
	Q9BTT0	Acidic leucine-rich nuclear phosphoprotein 32 family member E	158 - 203	EEEDDEDGDEDDEEEEENEAGPPEGYEEEEEEEEEEEDEDEDEDEDE	Histone binding, Phophatase inhibitor activity
	Q7Z6Z7	E3 ubiquitin-protein ligase HUWE1	2425 - 2469	EDEDDSQDEEEEEEEDEEDDQEDDEGEEGDEDDDDDGSEMELDED	DNA binding, ligase activity, poly(A) RNA binding
	Q12873-3	Isoform 3 of Chromodomain-helicase-DNA-binding protein 3	6-48	DEEEEEEEMVVSEEEEEEEEEGDEEEEEEVEAADEDDEEDDDE	DNA binding, Helicase activity, poly(A) RNA binding
	Q96KQ7	Histone-lysine N-methyltransferase EHMT2	289 - 331	EVEALTEQLSEEEEEEEEEEEEEEEEEEEEEEEEEEDEESGNQSD	Histone methyltransferase, p53 binding, C2H2 Zinc finger domain
	Q8IX15	Homeobox and leucine zipper protein Homez	507 - 549	EVVVCLDEEEEEEEEELPEDDEEEEEEEEEEDDDDDDDDVIIQD	DNA binding, Sequence-specific transcription factor, Transcription co-repressor
	Q8WYB5	Histone acetyltransferase KAT6B	1062 - 1103	ELSKESSEEEEEEEDEEEEEEEEEEEEEDEEEEEEEEEEEEEE	DNA binding, Histone acyltransferase, Transcription factor binding
	P19338	Nucleolin	233 - 274	EDEDEEEDDEDEDDDDDEDDEDDDDEDDEEEEEEEEEEPVKE	DNA binding, RNA binding, Protein binding
	Q5H9L4	Transcription initiation factor TFIID subunit 7-like	326 - 367	DSRSNNDDDEDEDDEDEDEDEDEDEDEDEDKEEEEEDCSEEYLE	Transcription coactivator, Transcription factor binding, Histone acyltransferase binding
	Q13029	PR domain zinc finger protein 2	261 - 301	EVNDLGEEEEEEEEEDEEEEEDDDDDELEDEGEEEASMPNE	DNA binding, sequence-specific DNA binding transcription factor, Zinc binding, histone-lysine N-methyltransferase
	P27797	Calreticulin	368 - 407	EEEEDKKRKEEEEAEDKEDDEDKDEDEEDEEDKEEDEEED	androgen receptor binding, carbohydrate binding, complement component C1q binding
	Q9UER7	DAXX_HUMAN Death domain-associated protein 6	433 - 471	ETDDEDDEESDEEEEEEEEEEEEEEEATDSEEEEDLEQMQE	Androgen receptor binding, Heat shock protein binding, Histone binding
	Q4LE39	AT-rich interactive domain-containing protein 4B	528 - 566	DETNKEEDEDDEEAEEEEEEEEEEEEEDEDDDDNNEEEEEFE	DNA binding, Protein binding, Transcription regulatory region DNA binding
	Q9UPS6-2	SET1B_HUMAN Isoform 2 of Histone-lysine N-methyltransferase SETD1B	1042 - 1079	EEQESTEEEEEAEEEEEEEEDDDDDDSDDRDESENDDED	Histone-lysine N-methyltransferase, Nuycleotide binding, RNA binding
	P39687	AN32A_HUMAN Acidic leucine-rich nuclear phosphoprotein 32 family member A	164 - 201	EGLDDEEEDEDEEEYDEDAQVVEDEEDEDEEEEGEEED	Gene expression, Intracellular signal transduction, nucleocytoplasmic transport
	P09429	High mobility group protein B1	178 - 214	EKSKKKKEEEEDEEDEEDEEEEEDEEDEDEEEDDDDE	DNA binding, Protein binding, Transcription factor binding
	Q9BT43	DNA-directed RNA polymerase III subunit RPC7-like	157 - 192	EEEVTSEEDEEKEEEEEKEEEEEEEEYDEEEHEEETD	Gene expression, Innate immune response, RNA polymerase III activity
	P17480	UBF1_HUMAN Nucleolar transcription factor 1	710 - 745	ESSSEDESEDGDENEEDDEDEDDDEDDDEDEDNESE	poly(A) RNA binding, RNA pol I CORE element seq-specific DNA binding, RNA pol I upstream control element seq-specific DNA binding
	Q15911	Zinc finger homeobox protein 3	453 - 488	EKVEPAEEEAEEEEEEEEEAEEEEEEEEEEEEEEEEDE	DNA binding, sequence-specific DNA binding transcription factor, Protein binding
	Q9UK99	F-box only protein 3	417 - 451	DEYEEMEEEEEEEEEEEDEDDDSADMDESDEDDEEE	ubiquitin-protein transferase activity
	Q9Y4B6	Protein VPRBP	1395 - 1429	EDEDEEEDQEEEEQEEEDDDEDDDDTDDLDELDTD	Histone kinase, Ser/Thr kinase, Protein binding
DNA-related (Replication, Repair)	P07199	Major centromere autoantigen B	403 - 446 504 - 537	EGEEEEEEEEEEEEEEEGEGEEEEEEEGEEEEEEEGGEGEELGEEEE EGGEDSDSDSEEEDDEEEDDEDEDDDDDEEDGDE	Centromeric DNA binding, Chromatin binding, DNA binding
	P20962	Parathymosin	38 - 74	EEEENGAEEEEEEETAEDGEEEDEGEEEDEEEEEEEDDE	DNA replication, Immune system process
RNA-related (Processing, Translation)	Q96MU7	YTH domain-containing protein 1	198 - 264	ENEEEGVEEDVEEDEEVEEDAEEDEEVDEDGEEEEEEEEEEEEEEEEEEEEYEQDERDQKEEGNDYD	poly(A) RNA binding, RNA binding
	O60841	Eukaryotic translation initiation factor 5B	528 - 566	ENPEEEEEEEEEEEEDEESEEEEEEEGESEGSEGDEEDE	GTPase activity, poly(A) RNA binding, GTP binding
	P12270	Nucleoprotein TPR	1948 - 1983	DDEEEDDDENDGEHEDYEEDEEDDDDDEDDTGMGDE	chromatin binding, heat shock protein binding, mRNA binding
	Q6ZU64	Coiled-coil domain-containing protein 108	1768 - 1803	EEEEEELEEEEEEEEETEEEELGKEEIEEKEEERDE	poly(A)RNA binding
	Q9NW13	RNA-binding protein 28	223 - 257	EEEDMEEEENDDDDDDDDEEDGVFDDEDEEEENIE	nucleotide binding, poly(A) RNA binding
	Q9UQ88	Cyclin-dependent kinase 11A	291 - 323	EEEEEEEEEEEEEGSTSEESEEEEEEEEEEEEE	ATP binding, cyclin-dependent protein ser/thr kinase
	P21127	Cyclin-dependent kinase 11B	303 - 335	EEEEEEEEEEEEEGSTSEESEEEEEEEEEEEEE	ATP binding, cyclin-dependent protein ser/thr kinase, poly(A) RNA binding
Other	Q5TCY1	Tau-tubulin kinase 1	732 - 779	EEEEEEEEDEEEEEEDEEEEEEEEEEEEEEEEEEEEEEEEAAAAVALGE	ATP binding, protein serine/threonine kinase activity
	P46060	Ran GTPase-activating protein 1	358 - 404	DDEDEEEEEEGEEEEEEAEEEEEEDEEEEEEEEEEEEEEPQQRGQGE	GTPase activator activity
	Q5JTC6	APC membrane recruitment protein 1	369 - 410	EEMALPDDDDEEEEEEEEVELEEEEEEVKEEEEDDDLEYLWE	beta-catenin binding, phosphatidylinositol-4,5-bisphosphate binding
	O60721	Sodium/potassium/calcium exchanger 1	854 - 894	DGGDSEEEEEEEEEQEEEEEEEEQEEEEEEEEEEEEKGNEE	calcium, potassium:sodium antiporter activity, symporter activity
	P21817	RYR1_HUMAN Ryanodine receptor 1	1872 - 1911	EEEEEEDEEEEGEEEDEEEKEEDEEETAQEKEDEEKEEEE	Calcium ion channel, Calmoduling binding
	O43847	NRDC_HUMAN Nardilysin	141 - 179	DDEEEEEVEEEEDDDEDSGAEIEDDDEEGFDDEDEFDDE	Epidermal growth factor binding, Metalloendopeptidase, Zinc ion binding
Function not clear	Q86TY3	Uncharacterized protein C14orf37	604 - 651	DQLESEEGQEDEDEEDEEDEDEEEEDEEEDEEDKDADSLDEGLDGDTE	Membrane
	Q7L0X2	Glutamate-rich protein 6	16 - 63	DQKESEEELEEEEEEEEVEEEEEEVEEEEEEVEEEEEEVVEEELVGEE	NA
	Q8TC90	Coiled-coil domain-containing glutamate-rich protein 1	301 - 344	EEEEEVEDEEEEVEDEEEEEVEEAEYVEEGEEELEEEELEEEEE	NA
	P0C7V8	DDB1- and CUL4-associated factor 8-like protein 2	107 - 146	EEETEREEEDEEIQEEGGEEEEEEEEEEEEEEEEEEEEEE	NA

Proteins are clustered into different categories depending on the biological process in which they are involved. Each protein is described by UniProt accession code (1st column), protein name (2nd column) and a list of GO terms (5th column). The corresponding acidic domains are described by their position in the coding sequence (3rd column) and their sequence (4th column).

Extended Data Table 2 | A list of human proteins containing KRDs with at least five lysines where three or more lysines are annotated as acetylation sites in the SSPKA database

	UniProt ID	Protein Name	Acetylated Lysines	Sequence of Lysine-rich Domain
Transcription Factor	O15525	Transcription factor MafG	53, 60, 71, 76	EEIVQLKQRRRTLKNRGYAASCRVKRVTQKEELEKQ
	P18146	Early growth response protein 1	422, 424, 425	KIHLRQKDKKADKSVV
	P52630	Signal transducer and activator of transcription 2	182, 184, 194, 197	RYKIQAKGKTPSLDPHQTKEQKILQETL
	Q16236	Nuclear factor erythroid 2-related factor 2	533, 536, 538, 541, 543, 548, 554, 555	QDLDHLKDEKEKLLKEKGENDKSLHLLKKQLSTLY
	Q9Y2Y9	Krueppel-like factor 13	166, 168, 180	LESPQRKHKCHYAGCEKVYGKSSHLKA
	P04150	Glucocorticoid receptor	480, 492, 494, 495	PACRYRKCLQAGMNLEARKTKKKIKGIQ
	P43694	Transcription factor GATA-4	312, 319, 321, 323	RPLAMRKEGIQTRKRKPKNLNKSK
	P06733*	Alpha-enolase	60, 71, 80, 89	KTRYMGKGVSKAVEHINKTIAPALVSKKLNVTEQEKIDKLMI
	P23769	Endothelial transcription factor GATA-2	389, 390, 399, 403, 405, 406, 408, 409	NRPLTMKKEGIQTRNRKMSNKSKKSKKGAECFE
Transcriptional Regulation (Except Transcription Factor), Chromatin Remodeling	O60563	Cyclin-T1	380, 386, 390	SQKQNSKSVPSAKVSLKEYRAKH
	P04406*	Glyceraldehyde-3-phosphate dehydrogenase	251, 254, 259, 260	LTCRLEKPAKYDDIKKVVKQAS
	P06748*	Nucleophosmin	141, 150, 154, 155 250, 257, 267, 273	LLSISGKRSAPGGGSKVPQKKVKLAAD VEDIKAKMQASIEKGGSLPKVEAKFINYVKNCFRMT
	P09874	Poly [ADP-ribose] polymerase 1	498, 505, 508	VVAPRGKSGAALSKKSKGQVKEE
	P19338	Nucleolin	70, 79, 87 102, 109, 116, 124, 132	VVVSPTKKVAVATPAKKAAVTPGKKAAATP KTVTPAKAVTTPGKKGATPGKALVATPGKKGAAIPAKGAKNGK
	P51531	Probable global transcription activator SNF2L2	996, 997, 999, 1003 1547, 1551, 1553, 1555, 1556	DGSEKDKKGKGGAKTLMNTI LNKKDDKGRDKGKGKKRPNRGK
	Q00987	E3 ubiquitin-protein ligase Mdm2	466, 467, 469, 470	ACFTCAKKLKKRNKPCP
	Q13547	Histone deacetylase 1	432, 438, 439, 441	EGEGGRKNSSNFKKAKRVKTED
	Q92793	CREB-binding protein	1797, 1806, 1809 1583, 1586, 1587, 1588, 1591, 1592, 1595, 1597	SLPSCQKMKRVVQHTKGCKRKTNGG GSQGDSKNAKKKNNKKTNKNKSSISRA
	Q92831	Histone acetyltransferase KAT2B	416, 428, 430, 441, 442	SSSPACKASSGLEANPGEKRKMTDSHVLEEAKKPRVMGD
	P27695*	DNA-(apurinic or apyrimidinic site) lyase	24, 27, 31, 32, 35	RTEPEAKKSKTAAKKNDKEAAGEG
	P62805	Histone H4	6, 9, 13, 17, 21, 32	MSGRGKGGKGLGKGGAKRHRKVLRDNIQGITKPAIRRL
	Q92922	SWI/SNF complex subunit SMARCC1	345, 346, 354, 359	SRKKSGKKGQASLYGKRRSQKEEDEQE
	P26358	DNA (cytosine-5)-methyltransferase 1	1111, 1113, 1115, 1117, 1119, 1121	SPGNKGKGKGKGKGKPKSQACEP
	Q13569	G/T mismatch-specific thymine DNA glycosylase	83, 84, 87	KKPVESKKSGKSAKSKE
	Q8TEK3	Histone-lysine N-methyltransferase, H3 lysine-79 specific	397, 398, 401	PSKARKKKLNKKGRKMA
	Q92841	Probable ATP-dependent RNA helicase DDX17	108, 109, 121, 129	GGGLPPKKFGNPGERLRKKKWDLSELPKFEKNFY
	P68431	Histone H3.1	5, 10, 15, 19, 24, 28, 37, 38	MARTKQTARKSTGGKAPRKQLATKAARKSAPATGGVKKPHRYRP
	Q92522	Histone H1x	179, 182, 185	KKGAGAKKDKGGKAKKTAA
	P46100	Transcriptional regulator ATRX	1933, 1935, 1936, 1939	YTKKKKKGKKGKKDSSSSG
	Q6DN03	Putative histone H2B type 2-C	13, 16, 17, 21, 24	FAPAPKKGSKKAVTKAQKKDGKKR
	P05114	Non-histone chromosomal protein HMG-14	3, 5, 14, 18, 27, 31, 38, 42, 48, 53, 55, 59, 61	MPKRKVSSAEGAAKEEPKRRSARLSAKPPAKVEAKPKKAAAKDKSSDKKVQTKGKRGAK GKQAEVAN
DNA Repair and Integrity	P12956	X-ray repair cross-complementing protein 6	539, 542, 544, 553, 556	DYNPEGKVTKRKHDNEGSGSKRPKVEYSEE
	Q9UQE7	Structural maintenance of chromosomes protein 3	105, 106, 113, 114	RRVIGAKKDQYFLDKKMVTKND
	P27695*	DNA-(apurinic or apyrimidinic site) lyase	24, 27, 31, 32, 35	RTEPEAKKSKTAAKKNDKEAAGEG
Other DNA Related Function	O94761	ATP-dependent DNA helicase Q4	376, 380, 382, 385, 386	RSRLLRKQAWKQKWRKKGECFGG
Ribosome Biogenesis	P06748*	Nucleophosmin	141, 150, 154, 155 250, 257, 267, 273	LLSISGKRSAPGGGSKVPQKKVKLAAD VEDIKAKMQASIEKGGSLPKVEAKFINYVKNCFRMT
Specific Molecular/Biological Function Uncertain	P81534	Beta-defensin 103	48, 54, 61, 66, 67	VLSCLPKEEQIGKCSTRGRKCCRRKK
	Q3BBV0	Neuroblastoma breakpoint family member 1	1101, 1103, 1105, 1106	VGEIEKKGKGKKRRGRRS
	Q8N7X0	Androglobin	337, 340, 343	KDGKEVKDVKEFKPESSLT
	Q6ZQR2	Uncharacterized protein C9orf171	237, 240, 246	EQKATQKAIKLEKKQKVVLGKL
Others	P04406*	Glyceraldehyde-3-phosphate dehydrogenase	251, 254, 259, 260	LTCRLEKPAKYDDIKKVVKQAS
	P09622	Dihydrolipoyl dehydrogenase, mitochondrial	267, 271, 273, 277	FQRILQKQGFKFKLNTKVTGATK
	P40939	Trifunctional enzyme subunit alpha, mitochondrial	350, 353, 359	HGQVLCKKNKFGAPQKDVKHLA
	Q9NP61	ADP-ribosylation factor GTPase-activating protein 3	223, 228, 229	KPNQAKKGLGAKKGSLGAQ
	Q9Y6F6	Protein MRVI1	398, 402, 405	EKRFAGKAGGKLAKAPGLKD
	P02768	Serum albumin	205, 214, 223, 229, 236 543, 548, 560, 565, 569, 581, 584, 588, 597, 598	AACLLPKLDELRDEGKASSAKQRLKCASLQKFGERAFKAWAVAR ICTLSEKERQIKKQTALVELVKHKPKATKEQLKAVMDDFAAFVEKCCKADDKETCFAEE GKKLVAASQ
	P62328	Thymosin beta-4	4, 12, 15	MSDKPDMAEIEKFDKSKLKKT
	Q13576	Ras GTPase-activating-like protein IQGAP2	1467, 1471, 1474	SIKLDGKGEPKGAKRAKPVK
	Q15283	Ras GTPase-activating protein 2	208, 209, 211	PSRNDQKKTKVKKKTS
	Q99075	Proheparin-binding EGF-like growth factor	96, 97, 99, 104	EHGKRKKKGKGLGKKRDPCLR
	P06733*	Alpha-enolase	60, 71, 80, 89	KTRYMGKGVSKAVEHINKTIAPALVSKKLNVTEQEKIDKLMI
	P15692	Vascular endothelial growth factor A	142, 147, 149, 152	RARQEKKSVRGKGKGQKRKRKKS
	P10636	Microtubule-associated protein tau	571, 574, 576, 584, 591, 597, 598, 607, 615	VPMPDLKNVKSKIGSTENLKHQPGGGKVQIINKKLDLSNVQSKCGSKDNIKHVPGGG

Each protein is described by its UniProt accession code and protein name (1st and 2nd column, respectively). Acetylated motifs are described by the position of their annotated acetylation sites within the coding sequence and their sequence (3rd and 4th column, respectively).

doi:10.1038/nature19754

Accessory subunits are integral for assembly and function of human mitochondrial complex I

David A. Stroud[1], Elliot E. Surgenor[1], Luke E. Formosa[1,2], Boris Reljic[2]†, Ann E. Frazier[3,4], Marris G. Dibley[1], Laura D. Osellame[1], Tegan Stait[3], Traude H. Beilharz[1], David R. Thorburn[3,4,5], Agus Salim[6] & Michael T. Ryan[1]

Complex I (NADH:ubiquinone oxidoreductase) is the first enzyme of the mitochondrial respiratory chain and is composed of 45 subunits in humans, making it one of the largest known multi-subunit membrane protein complexes[1]. Complex I exists in supercomplex forms with respiratory chain complexes III and IV, which are together required for the generation of a transmembrane proton gradient used for the synthesis of ATP[2]. Complex I is also a major source of damaging reactive oxygen species and its dysfunction is associated with mitochondrial disease, Parkinson's disease and ageing[3–5]. Bacterial and human complex I share 14 core subunits that are essential for enzymatic function; however, the role and necessity of the remaining 31 human accessory subunits is unclear[1,6]. The incorporation of accessory subunits into the complex increases the cellular energetic cost and has necessitated the involvement of numerous assembly factors for complex I biogenesis. Here we use gene editing to generate human knockout cell lines for each accessory subunit. We show that 25 subunits are strictly required for assembly of a functional complex and 1 subunit is essential for cell viability. Quantitative proteomic analysis of cell lines revealed that loss of each subunit affects the stability of other subunits residing in the same structural module. Analysis of proteomic changes after the loss of specific modules revealed that ATP5SL and DMAC1 are required for assembly of the distal portion of the complex I membrane arm. Our results demonstrate the broad importance of accessory subunits in the structure and function of human complex I. Coupling gene-editing technology with proteomics represents a powerful tool for dissecting large multi-subunit complexes and enables the study of complex dysfunction at a cellular level.

Mitochondrial complex I is a boot-shaped structure of ~1 MDa with a hydrophilic matrix arm and a hydrophobic membrane arm[7–9]. These arms are assembled via intermediate modules through transient association with assembly factors[10]. The N-module at the tip of the matrix arm is involved in the oxidation of NADH, whereas the Q-module bridges the matrix and membrane arms and is involved in transfer of electrons along Fe-S clusters to ubiquinone. With the reduction of ubiquinone, four protons are pumped across the inner membrane into the intermembrane space. The core structure of the membrane arm is defined by 7 subunits encoded by mitochondrial DNA (mtDNA); ND1 at the base of the Q-module, followed by ND3, ND6 and ND4L, and the antiporter-like subunits ND2, ND4 and ND5 (refs 8, 9). The mechanisms of NADH oxidation and proton pumping are conserved from bacteria to humans, with 14 core (including the 7 mtDNA-encoded) subunits performing these roles[11].

To investigate the importance of the 31 accessory subunits, we used TALEN and CRISPR/Cas9 gene-editing tools to disrupt their genes in human HEK293T cells (Supplementary Table 1). Of the knockout lines generated, 24 were unable to grow on galactose-containing media indicating mitochondrial respiration defects (Fig. 1). Blue-native (BN)–PAGE and immunoblot analysis for subunits NDUFA9, NDUFA13 and NDUFB11 (located in different regions of the complex) revealed that loss of an individual accessory subunit often disrupted assembly of complex I (Fig. 1). Analysis of the supercomplex was also disrupted in the same cell lines (Extended Data Fig. 1) whereas assembly of complexes III and IV was not affected (Extended Data Fig. 2). For cell lines still capable of growth on galactose, a complex was present that did not markedly differ from the migration of mature complex I (Fig. 1, lanes 2–6). Other cell lines showed different subcomplexes including one that migrated slightly faster than complex I, consistent with loss of the N-module[12] (Fig. 1, lanes 5–8, marked with ‡).

In contrast to all other subunits, we found NDUFAB1 to be essential for cell viability (Extended Data Fig. 3a–c). NDUFAB1 is unique as it is the only subunit with a 2:1 stoichiometry within the complex, where it binds LYR motifs present in NDUFA6 and NDUFB9 (ref. 7). NDUFAB1 is also the mitochondrial acyl carrier protein[13] and associates with proteins involved in fatty-acid synthesis (LIPT2) and other proteins (Extended Data Fig. 3d, Supplementary Table 2) including LYRM7 that promotes biogenesis of the complex III Rieske subunit (UQCRFS1). An NDUFAB1-knockout cell line was generated by complementing cells with the yeast mitochondrial acyl carrier protein (yACP1) (Extended Data Fig. 3e, Supplementary Table 1). Since the NDUFAB1-knockout lacks assembled complex I and dies in galactose media (Fig. 1, Extended Data Fig. 3c, e), the essential role of NDUFAB1 is independent of complex I.

We selected a subset of representative knockout lines for further analysis (Fig. 2a). Rescue of each line restored complex I assembly in all cases (Fig. 2b). Cell lines lacking NDUFV3, NDUFA12 or NDUFA7 that still grew on galactose had negligible to moderate reductions in complex I activity and mitochondrial respiratory capacity (Fig. 2c). Knockout of NDUFS6 led to most of the N-module dissociating from complex I; however, this did not severely affect complex I activity or respiratory capacity (Fig. 2c). This is consistent with previous patient cell studies[12] and suggests that the complex is less stable during BN–PAGE (Extended Data Fig. 4a). In NDUFA12-knockout cells, the complex I assembly factor NDUFAF2 substituted for its paralogue NDUFA12, leading to complex I appearing fully assembled (Extended Data Fig. 4b). In NDUFA2-knockout cells, no N-module was present (Extended Data Fig. 4a) and these cells showed severe defects in complex I activity and respiration (Fig. 2c). We propose that NDUFV3 may be the terminally assembled subunit of complex I owing to its location[9] and lack of defects upon its loss. *In vitro* imported NDUFV3 also readily exchanged with the endogenous assembled protein (Extended Data Fig. 4c), while only bona fide subunits were enriched without assembly factors when complex I was isolated using NDUFV3 as bait (Extended Data Fig. 3d, Supplementary Tables 3, 4). In contrast to most N-module subunits, knockout of membrane arm subunits

[1]Department of Biochemistry and Molecular Biology, Monash Biomedicine Discovery Institute, Monash University, 3800, Melbourne, Australia. [2]Department of Biochemistry and Genetics, La Trobe Institute for Molecular Science, La Trobe University 3086, Melbourne, Australia. [3]Murdoch Childrens Research Institute, Royal Children's Hospital, Melbourne 3052, Australia. [4]Department of Pediatrics, University of Melbourne, Melbourne 3052, Australia. [5]Victorian Clinical Genetics Services, Royal Children's Hospital 3052, Melbourne, Australia. [6]Department of Mathematics and Statistics, La Trobe University 3086, Melbourne Australia. †Present address: Walter and Eliza Hall Institute of Medical Research, Parkville, Melbourne, Victoria 3052, Australia.

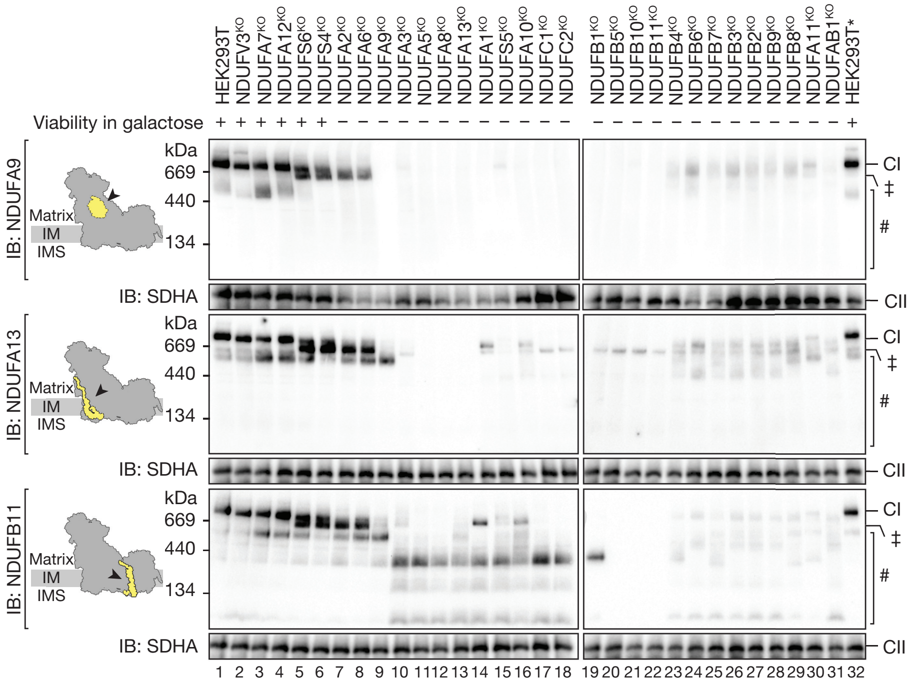


Figure 1 | Analysis of complex I assembly in knockout cell lines. Mitochondria were solubilized in Triton X-100 and analysed by BN–PAGE and immunoblotting (IB). CI, complex I; KO, knockout; IM, inner membrane; IMS, intermembrane space. SDHA (complex II subunit, CII) was used as a loading control. ‡, loss of N-module; #, subcomplexes. HEK293T* denotes control for NDUFAB1-knockout (NDUFAB1^{KO}). Galactose growth phenotypes and subunit positions indicated.

resulted in severe mitochondrial respiration defects (Fig. 2c) and loss of assembled complex I with the concomitant accumulation of subcomplexes (Fig. 1).

The severity of knockouts observed for each accessory subunit on complex I assembly seemed to largely predict the effect of mutations in patients with mitochondrial disease[14] (Extended Data Table 1). Almost all patients with mutations in genes encoding three of the subunits that exhibit mild assembly defects (NDUFA12, NDUFS4 and NDUFS6) have two nonsense mutations that block subunit expression. By contrast, patients with defects in 8 accessory subunits, showing severe assembly defects, carry missense mutations, suggesting that the complete loss of any of these subunits may be incompatible with human life (Extended Data Table 1).

Next we used stable isotope labelling with amino acids in cell culture (SILAC) and quantitative mass spectrometry to determine changes in levels of cellular proteins in the representative knockout lines (Supplementary Table 5). Most of the >6,000 cellular proteins detected did not significantly differ from control except for complex I subunits themselves, which were consistently downregulated (Fig. 2e, Extended Data Fig. 5a). Of the other 20 mitochondrial proteins

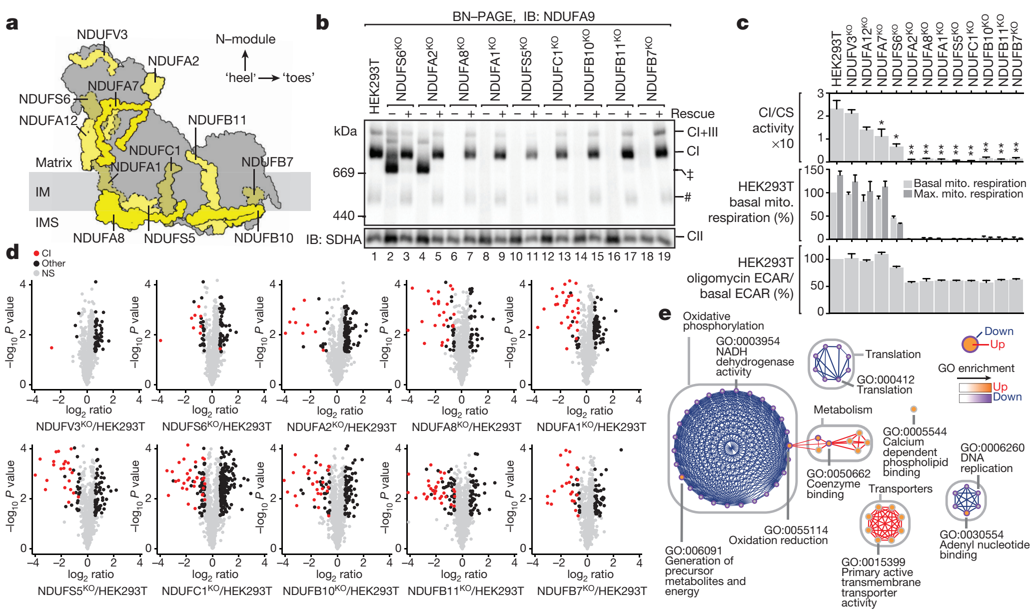


Figure 2 | Metabolic and proteomic analysis of representative complex I accessory subunit-knockout lines. a, Positions of subunits in complex I (ref. 9). **b**, Cell lines complemented with cDNA encoding the targeted gene. Analysis as in Fig. 1. CI+III, supercomplex; ‡, loss of N-module; #, subcomplexes. **c**, Top, complex I (CI) activity relative to citrate synthase (CS). $n = 3$ or 4 (HEK293T, NDUFA7 and NDUFV3) biological replicates. $^{*}P < 0.05$; $^{**}P < 0.01$, unpaired t-test. Middle, mitochondrial basal and maximal respiration rates. Bottom, glycolytic capacity. ECAR, extracellular acidification rate. Middle and bottom panels, $n = 3$ or 4 (NDUFA7, NDUFA12 and NDUFV3) biological replicates. Data are mean ± s.e.m. **d**, Volcano plots showing relative levels of proteins in knockout cells. $n = 3$ biological replicates. Red dots, complex I subunits. Black dots, $P < 0.05$, >1.5-fold change, unpaired t-test. Light grey dots, not significant (NS) ($P > 0.05$, <1.5-fold change). **e**, Gene Ontology (GO) enrichment map of pathways and functions altered in respiration deficient knockouts. Example GO terms are grouped according to general role.

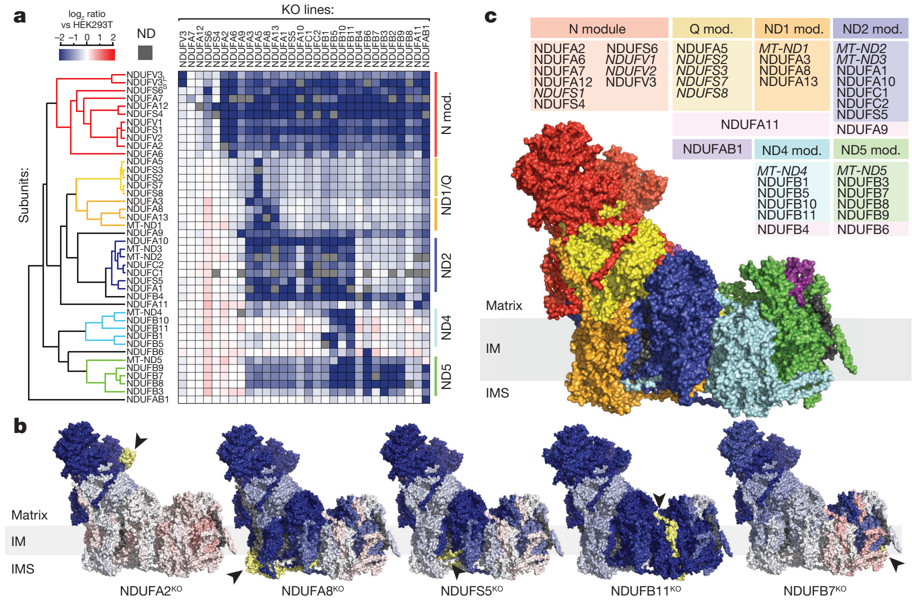


Figure 3 | Subunit stability correlates with structural modules. **a**, Levels of complex I subunits in knockout lines. ND, not detected. **b**, Subunit levels for knockouts mapped to the complex I structure[9]. Grey, no data; yellow and arrow, knockout subunit. Scale as in **a**. **c**, Clusters defined in **a** mapped to the complex I structure. Italics denote core subunits; subunits not clustered have been removed for clarity.

in which levels were changed more than 2-fold, 9 were similarly responsive in a cell line lacking functional complex IV (ref. 15; Extended Data Fig. 5b, Supplementary Table 6), pointing to these gene products being related to general defects in oxidative phosphorylation. Besides affected gene sets related to complex I and oxidative phosphorylation, other affected pathways were related to metabolism, transporter activity, translation and DNA replication (Fig. 2e, Extended Data Fig. 5d).

Hierarchical clustering of protein ratios in the representative knockout cell lines (Extended Data Fig. 5a) identified clusters of complex I subunits that are similarly located in the structure (for example, NDUFS4, NDUFA7 and NDUFA12). To increase the resolution of our clustering analysis, we measured the levels of mitochondrial proteins in the remaining 20 knockout cell lines (Fig. 3a, Supplementary Table 5). Using the colour scheme from our heat maps, we mapped the levels of individual subunits in each knockout onto the recently solved structure of bovine complex I (ref. 9; Fig. 3b, Extended Data Fig. 6a). We uncovered clear structural correlations including the loss of subunits around the N-module upon knockout of subunit NDUFA2 as well as loss of subunits from the distal membrane module in NDUFB11-knockout cells (Fig. 3b). Transcriptomic analysis of the representative knockout cell lines revealed that the only genes with more than a twofold difference in expression are those that were gene-edited (Extended Data Fig. 7). We conclude that the mutated target genes may be subject to nonsense-mediated mRNA decay while the other complex I subunits in which levels decrease are most likely proteolytically degraded[16].

Hierarchical clustering analysis of complex I subunits (Supplementary Table 7) identified five clusters containing subunits with similar stabilities across knockouts (Fig. 3a). Mapping of these clusters to the structure of bovine complex I (ref. 9) revealed distinct modules (Fig. 3c, Extended Data Fig. 6b). One cluster contains subunits encompassing the N-module while every other cluster partitions with mtDNA-encoded 'ND' subunits. NDUFAB1 could not be assigned to any cluster, with its level being almost unchanged, consistent with its separate functions. Subunits NDUFA9, NDUFB4, NDUFB6 and NDUFA11 were not clearly mapped to an individual module and may reside at module interfaces.

Complex I is assembled via a series of intermediate assembly modules and requires the involvement of >10 known assembly factors[17,18]. We generated knockout cell lines of assembly factors known to function at different steps—NDUFAF1, NDUFAF2, NDUFAF4, NDUFAF6 and TIMMDC1 (Extended Data Fig. 8a). BN–PAGE analysis showed a reduction or loss of complex I assembly (Fig. 4a). Proteomic analysis indicated that complex I subunits belonging to different modules were affected to varying degrees (Extended Data Fig. 8b, c). The profile of changes in complex I subunits in assembly-factor-knockout lines correlated with groups of complex I subunit-knockout lines in which the subunits belong to distinct modules (Fig. 4b) consistent with assembly models[18]. Since little is known about the assembly of the distal membrane module, we searched our proteomic data set for proteins altered in knockouts belonging to the ND4 and ND5 module relative to those belonging to ND1 and ND2 modules (Supplementary Table 8). ATP5SL, recently identified in a complex I subassembly[17], accumulated in ND4- and ND5-module knockout lines (Fig. 4c). In a separate analysis, the uncharacterized TMEM261, which we later termed DMAC1 (see below), was at increased levels in membrane arm subunit knockout lines when compared against matrix arm subunit knockout lines (Fig. 4c, Supplementary Table 9). Knockout of either ATP5SL or DMAC1 led to specific and severe complex I assembly defects (Fig. 4d, Extended Data Fig. 9a, b) and turnover of N-module and distal membrane arm subunits (Extended Data Fig. 9c, d). Integration of the proteomic profiles in DMAC1- and ATP5SL-knockout lines with those originating from our accessory subunit knockouts indicated a strong correlation with the ND5 module (Fig. 4b).

While DMAC1 is absent from the MitoCarta2.0 database[19], we found it to be a mitochondrial inner-membrane protein (Extended Data Fig. 9e, f). Pulse-chase analysis revealed that mtDNA-encoded subunits formed a 600-kDa intermediate complex[20] in DMAC1-knockout cells but then dissociated (Extended Data Fig. 9g), indicating a late-stage assembly defect similar to that seen upon loss of complex I assembly factor FOXRED1 (ref. 21). Proteins highly enriched with ATP5SL included complex I subunits of the ND4 module and FOXRED1 (Fig. 4e), whereas proteins enriched with DMAC1 included subunits ND4 and ND5, plus ATP5SL and FOXRED1 along with OXA1L, the membrane insertase for mtDNA-encoded subunits (Fig. 4e, Supplementary Tables 10, 11). ATP5SL and DMAC1 also interacted with newly translated ND5 (Fig. 4f). Since other complex I subunits, assembly factors and subunits of complexes III and IV were enriched in DMAC1 pull-downs, the integration of the ND4 and ND5 modules

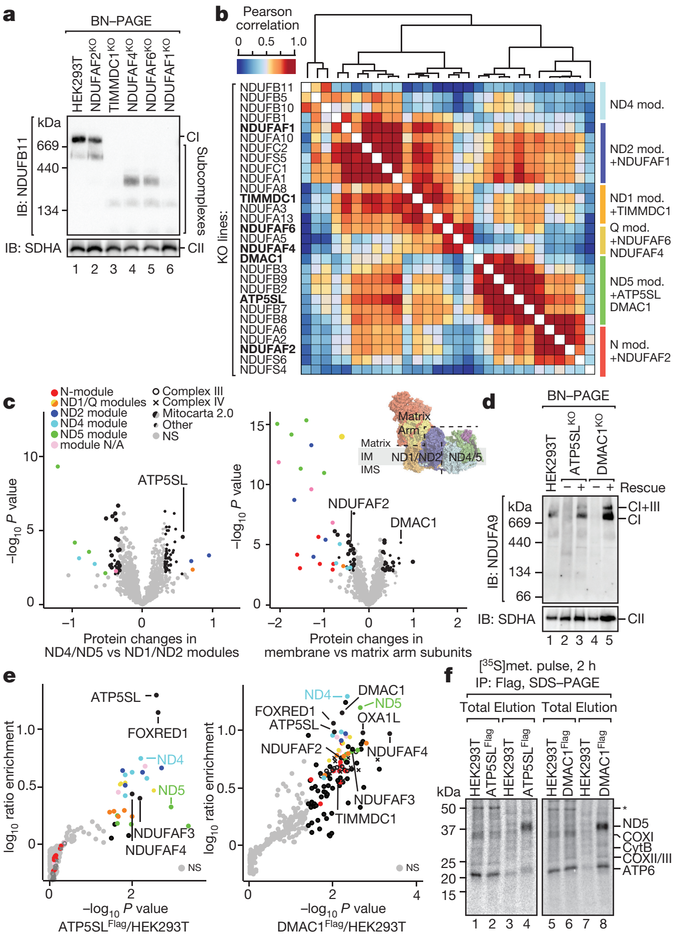


Figure 4 | Analysis of complex I assembly factors including DMAC1 and ATP5SL. a, Complex I in assembly factor-knockout lines as per Fig. 1. **b**, Knockout lines compared via Pearson correlation and hierarchical clustering. **c**, Volcano plots showing proteins with altered levels of subunits in specific modules in knockouts. Light grey dots, not significant (>5% false discovery rate (FDR), <1.5-fold change). N/A, not assigned. *P* values from an unpaired *t*-test using permutation based FDR statistics; $n = 27$ biological replicates (ND4/ND5, ND1/ND2 modules), $n = 56$ (membrane arm), $n = 23$ (matrix arm). **d**, Complementation of knockouts as in Fig. 2b. **e**, Affinity enrichment of proteins from DMAC1$^{\text{Flag}}$ or ATP5SL$^{\text{Flag}}$ cells. *P* values are from an unpaired single-sided *t*-test. $n = 3$ biological replicates; light grey dots, not significant ($P > 0.05$). **f**, Radiolabelled mtDNA-encoded subunits in ATP5SL$^{\text{Flag}}$ or DMAC1$^{\text{Flag}}$ lines were immunoprecipitated with Flag beads, analysed by SDS–PAGE and autoradiography.

in the assembly pathway may intersect with supercomplex formation and occur concurrently with addition of the N-module, the final step in complex I assembly[22]. Owing to the association of DMAC1 with the biogenesis of the distal region of complex I, we termed the protein distal membrane-arm assembly component 1.

In summary, we demonstrate that accessory subunits are integrally associated in modules, defined by the core structural and functional subunits of human complex I, assembly of which require the concerted action of assembly factors. By defining the impact of individual subunit knockouts, our data will facilitate validation of putative pathogenic variants found in complex I genes in patients, while *DMAC1* and *ATP5SL* also represent new pathological gene targets. Our approach additionally serves as a powerful example of how coupling gene editing and quantitative proteomics allows rapid insights into previously inaccessible aspects of human cellular function.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 10 February; accepted 22 August 2016.
Published online 14 September 2016.

1. Sazanov, L. A. A giant molecular proton pump: structure and mechanism of respiratory complex I. *Nat. Rev. Mol. Cell Biol.* **16,** 375–388 (2015).
2. Lapuente-Brun, E. *et al.* Supercomplex assembly determines electron flux in the mitochondrial electron transport chain. *Science* **340,** 1567–1570 (2013).
3. Vafai, S. B. & Mootha, V. K. Mitochondrial disorders as windows into an ancient organelle. *Nature* **491,** 374–383 (2012).
4. Morais, V. A. *et al.* PINK1 loss-of-function mutations affect mitochondrial complex I activity via NdufA10 ubiquinone uncoupling. *Science* **344,** 203–207 (2014).
5. Miwa, S. *et al.* Low abundance of the matrix arm of complex I in mitochondria predicts longevity in mice. *Nat. Commun.* **5,** 3837 (2014).
6. Hirst, J. Mitochondrial complex I. *Annu. Rev. Biochem.* **82,** 551–575 (2013).
7. Vinothkumar, K. R., Zhu, J. & Hirst, J. Architecture of mammalian respiratory complex I. *Nature* **515,** 80–84 (2014).
8. Zickermann, V. *et al.* Structural biology. Mechanistic insight from the crystal structure of mitochondrial complex I. *Science* **347,** 44–49 (2015).
9. Zhu, J., Vinothkumar, K. R. & Hirst, J. Structure of mammalian respiratory complex I. *Nature* **536,** 354–358 (2016).
10. Pagliarini, D. J. *et al.* A mitochondrial protein compendium elucidates complex I disease biology. *Cell* **134,** 112–123 (2008).
11. Baradaran, R., Berrisford, J. M., Minhas, G. S. & Sazanov, L. A. Crystal structure of the entire respiratory complex I. *Nature* **494,** 443–448 (2013).
12. Lazarou, M., McKenzie, M., Ohtake, A., Thorburn, D. R. & Ryan, M. T. Analysis of the assembly profiles for mitochondrial- and nuclear-DNA-encoded subunits into complex I. *Mol. Cell. Biol.* **27,** 4228–4237 (2007).
13. Runswick, M. J., Fearnley, I. M., Skehel, J. M. & Walker, J. E. Presence of an acyl carrier protein in NADH:ubiquinone oxidoreductase from bovine heart mitochondria. *FEBS Lett.* **286,** 121–124 (1991).
14. Rodenburg, R. J. Mitochondrial complex I-linked disease. *Biochim. Biophys. Acta* **1857,** 938–945 (2016).
15. Stroud, D. A. *et al.* COA6 is a mitochondrial complex IV assembly factor critical for biogenesis of mtDNA-encoded COX2. *Hum. Mol. Genet.* **24,** 5404–5415 (2015).
16. Quirós, P. M., Langer, T. & López-Otín, C. New roles for mitochondrial proteases in health, ageing and disease. *Nat. Rev. Mol. Cell Biol.* **16,** 345–359 (2015).
17. Andrews, B., Carroll, J., Ding, S., Fearnley, I. M. & Walker, J. E. Assembly factors for the membrane arm of human complex I. *Proc. Natl Acad. Sci. USA* **110,** 18934–18939 (2013).
18. Sánchez-Caballero, L., Guerrero-Castillo, S. & Nijtmans, L. Unraveling the complexity of mitochondrial complex I assembly: A dynamic process. *Biochim. Biophys. Acta* **1857,** 980–990 (2016).
19. Calvo, S. E., Clauser, K. R. & Mootha, V. K. MitoCarta2.0: an updated inventory of mammalian mitochondrial proteins. *Nucleic Acids Res.* **44,** D1251–D1257 (2015).
20. Stroud, D. A., Formosa, L. E., Wijeyeratne, X. W., Nguyen, T. N. & Ryan, M. T. Gene knockout using transcription activator-like effector nucleases (TALENs) reveals that human NDUFA9 protein is essential for stabilizing the junction between membrane and matrix arms of complex I. *J. Biol. Chem.* **288,** 1685–1690 (2013).
21. Formosa, L. E. *et al.* Characterization of mitochondrial FOXRED1 in the assembly of respiratory chain complex I. *Hum. Mol. Genet.* **24,** 2952–2965 (2015).
22. Mimaki, M., Wang, X., McKenzie, M., Thorburn, D. R. & Ryan, M. T. Understanding mitochondrial complex I assembly in health and disease. *Biochim. Biophys. Acta* **1817,** 851–862 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank M. Curtis, P. Faou, M. Lazarou, B. Porebski, L. Twigg, R. Schittenhelm (Monash Biomedical Proteomics Platform), A. Barugahare and P. Harrison (Monash Bioinformatics Platform), Monash Micro Imaging and the Micromon NGS Facility for assistance. We acknowledge funding from NHMRC Project Grants (1068056, 1107094) and fellowships (1070916 to D.A.S., 541920 to A.E.F., 1022896 to D.R.T.), the Australian Mitochondrial Disease Foundation and the Victorian Government's Operational Infrastructure Support Program.

Author Contributions D.A.S. and M.T.R. conceived the project and wrote the manuscript; D.A.S., D.R.T. and M.T.R. designed the experiments; D.A.S., E.E.S., L.E.F., B.R., M.G.D., L.D.O. and M.T.R. generated and analysed knockout lines; D.A.S. performed proteomic experiments; A.E.F. and T.S. performed enzymology; T.H.B. undertook transcript analysis; A.S. developed normalization algorithms.

Author Information Data are available via ProteomeXchange under accession PXD004666, and the NCBI Gene Expression Omnibus (GEO) under accession GSE84913. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.A.S. (d.stroud@monash.edu) or M.T.R. (michael.ryan@monash.edu).

Reviewer Information *Nature* thanks J. Hirst, B. Lightowlers and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Cell lines, gene editing and screening. HEK293T cells[23], commonly used in complex I assembly studies[15,20,21,24,25], interactome[26] and mitochondrial complexome studies[24], were originally purchased from the ATCC and a clonal cell line was obtained after single cell sorting[20] and used as the parental line for all gene editing and proteomic work. Knockout cell lines were validated by sequencing of targeted alleles for insertions and deletions (indels), immunoblotting and subsequent proteomic analysis. Cell lines regularly undergo testing for mycoplasma contamination using PlasmaTest (InvivoGen). Gene editing was performed using TALEN[27] pairs as described[15,28], or the pSpCas9(BB)-2A-GFP (PX458) CRISPR/Cas9 construct (a gift from F. Zhang; Addgene, plasmid 48138; ref. 29). In brief, in the first round, TALEN constructs were designed using the ZiFiT Targeter[30]. For genes unsuccessfully targeted in the first round, CRISPR/Cas9 guide RNAs were designed for a second round of gene-disruption using CHOPCHOP[31]. Successful targeting strategies and constructs can be found in Supplementary Table 1. Gene edited and control HEK293T cells[15] were cultured in DMEM (ThermoFisher) supplemented with 10% (v/v) FBS and 50 μg ml^{-1} uridine. Transfection reagents used were Lipofectamine 2000 and Lipofectamine LTX (ThermoFisher). During screening, glucose-free DMEM supplemented with 5 mM galactose, 1 mM sodium pyruvate, 10% (v/v) dialysed FBS (ThermoFisher) and 50 μg ml^{-1} uridine was used to identify respiratory incompetent knockout clones. Respiratory competent knockout clones were identified by sequencing of a mixed PCR product covering the target region, where a loss of sequencing fidelity at the target indicates a candidate clone[28]. With the exception of the NDUFA9- and COA6-knockout cell lines, which were described previously[15,20], indels for individual alleles are summarized in Supplementary Table 1.

To generate NDUFAB1 knockout cells, clonal HEK293T cells were transduced with lentiviruses pLVX-TetOne-Puro-NDUFAB1^{*Flag} or pLVX-TetOne-Puro-yACP1^{Flag} (Clontech). NDUFAB1^{*Flag} represents the C-terminally Flag-tagged human NDUFAB1 protein encoded by cDNA having undergone silent mutagenesis to remove the CRISPR/Cas9 target site. yACP1^{Flag} indicates cDNA encoding the C-terminally Flag-tagged yeast (*Saccharomyces cerevisiae*) ACP1. Transduced cells were grown in the presence of 2 μg ml^{-1} puromycin for 72 h, and expression of NDUFAB1^{*Flag} or yACP1^{Flag} was confirmed after a further 72 h of treatment with 1 μg ml^{-1} doxycycline (DOX; Sigma-Aldrich) followed by SDS–PAGE and immunoblotting with NDUFAB1 (Abcam) and Flag (Sigma-Aldrich) antibodies. For subsequent gene editing, cells cultured in the presence of 50 ng ml^{-1} DOX were transfected with pSpCas9(BB)-2A-GFP-NDUFAB1 and screened as described above.

For complementation, cDNAs encoding NDUFV3^{Flag}, NDUFS6^{Flag}, NDUFA8^{Flag}, ATP5SL^{Flag} and DMAC1^{Flag} (TMEM261^{Flag}) were cloned into pBABE-puro (Addgene, 1764; ref. 32), whereas NDUFA1, NDUFA2, NDUFB7, NDUFB10, NDUFB11 and NDUFC1 cDNAs were cloned into pBMN-Z (Addgene, 1734) in place of the LacZ insert. Retroviral constructs were used to transduce the corresponding main clone (Supplementary Table 1), following which expression was selected for through growth in galactose DMEM with the exception of NDUFS6 and NDUFV3 knockouts which were selected using 2 μg ml^{-1} puromycin. Transduction was verified by BN–PAGE or SDS–PAGE followed by immunoblotting with NDUFA9 or Flag antibodies, respectively.

Mitochondrial isolation, gel electrophoresis, immunoblotting and antibodies. Mitochondria were isolated as previously described[33]. Protein concentration was estimated by bicinchoninic acid assay (BCA; Pierce), and aliquots of crude mitochondria stored at −80 °C until use. SDS–PAGE was performed using samples solubilized in LDS sample buffer and separated on NuPAGE Novex Bis-Tris protein gels according to manufacturer's instructions (ThermoFisher). Tris-Tricine SDS–PAGE, BN–PAGE and 2D–PAGE were performed as described previously[34–36]. Carbonate and swelling experiments were performed as described[37]. Immunoblotting onto PVDF membranes was performed using a Novex Semi-Dry Blotter (ThermoFisher) according to manufacturer's instructions. Horseradish peroxidase coupled secondary antibodies and ECL chemiluminescent substrate (BioRad) were used for detection on a BioRad ChemiDoc XRS+ imaging system. The following primary antibodies were used in this study: COX2 (ThermoFisher A-6404), COX4 (Abcam, ab110261), Flag (Sigma-Aldrich, M2 clone), MIC10 (Aviva Systems Biology, ARP44801_P050), NDUFA13 (Mitosciences MS103-SP), NDUFAB1 (Abcam, ab96230), NDUFB11 (Abcam, ab183716), NDUFV1 (Proteintech 11238-1-AP), NDUFS2 (Mitosciences, MS114), anti-respiratory-chain (Abcam, ab110413; which contains antibodies against ATP5A, UQCRC2, COX1, SDHB and NDUFB8), SDHA (Abcam, ab14715), TIMMDC1 (Sigma, HPA053214), TOMM20 (Santa Cruz, Sc11415) and UQCRC1 (ThermoFisher, 16D10AD9AH5), while rabbit polyclonal antibodies against NDUFA9 (ref. 12), NDUFAF1 (also known as CIA30)[38], NDUFAF2 (ref. 21), NDUFAF4 (ref. 21), NDUFB6 (ref. 38) and HSP70 (ref. 20) were raised in-house.

mRNA expression level analysis. For analysis of mRNA expression levels, total RNA was obtained from each cell line in replicate with TRIzol (Thermo scientific). Total RNA was purified using Direct-zol columns according to the manufacturer's specifications (Zymo Research). For cDNA synthesis, 1 μg of total RNA was processed as the T12VN-PAT assay[39] adapted for multiplexing on the Illumina MiSeq instrument. We refer to this assay as mPAT for multiplexed PAT. The approach is based on a nested PCR that sequentially incorporates the Illumina platform's flow-cell-specific terminal extensions onto 3′ RACE PCR amplicons. First, cDNA was generated using the anchor primer mPAT Reverse, next this primer and a pool of 50 gene-specific primers were used in 5 cycles of amplification. Each gene-specific primer had a universal 5′ extension (see Supplementary Table 12) for sequential addition of the 5′ (P5) Illumina elements. These amplicons were then purified using NucleoSpin columns (Macherey-Nagel), and entered into second round of amplification using the universal Illumina Rd1 sequencing Primer and TruSeq indexed reverse primers from Illumina. Second-round amplification was for 14 cycles. Note, that each experimental condition was amplified separately in the first round with identical primers. In the second round, a different indexing primer was used for each experimental condition. All PCR reactions were pooled and run using the MiSeq Reagent Kit v2 with 300 cycles (that is, 300 bases of sequencing) according to the manufacturer's specifications. Data were analysed using established bioinformatics pipelines[40]. Figures were generated using the R framework.

Oxygen consumption and enzymatic activity measurements. Oxygen consumption (OCR) and extracellular acidification (ECAR) rates were measured in live cells using a Seahorse Bioscience XF24-3 Analyzer as described[15]. In brief, 50,000 cells were plated per well in Seahorse Bioscience culture plates treated with 50 μg ml poly-D-lysine and grown overnight in standard culture conditions. The cellular OCR and ECAR were analysed in non-buffered DMEM (Seahorse Biosciences) containing 5 mM glucose, 1 mM sodium pyruvate and 50 μg ml^{-1} uridine with the following inhibitors: 2 μM oligomycin; 0.5 μM carbonyl cyanide 4-(trifluoromethoxy) phenylhydrazone (FCCP); 0.5 μM rotenone; and 0.3 μM antimycin A. For each assay cycle, four measurement time points of 2 min mix, 2 min wait and 5 min measure were collected. For each cell line, 3–4 replicate wells were measured in multiple plates and CyQuant (ThermoFisher Scientific) was used to normalize measurements to cell number. Basal OCR and non-mitochondrial respiration (following rotenone and antimycin A injections) were calculated as a mean of the measurement points. Basal ECAR was calculated from the initial basal measurement cycle. To calculate proton-leak and maximal respiration, the initial measurement following addition of oligomycin or FCCP was used. Enzymatic activity measurements were performed as previously described[41] in three separate subcultures of each cell line. To accommodate unequal variance, statistical significance was determined through an unpaired two-sample, two-sided *t*-test using Welch's correction.

Radiolabelling of mtDNA-encoded translation products and protein import. Radiolabelling of mtDNA-encoded proteins was performed as previously described[15,34]. Isolated mitochondria were subjected to BN–PAGE or 2D–PAGE as described above, following which proteins were transferred to PVDF membranes and analysed by phosphorimager digital autoradiography (GE Healthcare Life Sciences). For immunoprecipitation of newly translated proteins, mitochondria were isolated from cells pulsed for 2 h and solubilized in 1% (w/v) digitonin, 20 mM Bis-Tris (pH 7.0), 50 mM NaCl, 0.1 mM EDTA, 10% (v/v) glycerol. After a brief clarification spin, complexes were incubated with anti-Flag affinity gel (SigmaAldrich), the gel washed with 0.2% (w/v) digitonin, 20 mM Bis-Tris (pH 7.0), 60 mM NaCl, 0.5 mM EDTA, 10% (v/v) glycerol, and enriched proteins eluted with the addition of 150 μg ml^{-1} Flag peptide (SigmaAldrich). Samples were TCA precipitated to remove detergent and analysed by SDS–PAGE and phosphorimaging as above.

For protein import, *NDUFA12*, *NDUFA7* and *NDUFV3* cDNA was cloned into the pGEM-4Z plasmid (Promega). mRNA was transcribed using the mMESSAGE mMACHINE SP6 transcription kit (ThermoFisher Scientific) according to the manufacturer's instructions. Radiolabelled proteins were translated in the presence of [^{35}S]methionine/cysteine using a rabbit reticulocyte lysate system (Promega). Translated proteins were incubated with isolated mitochondria at 37 °C as previously described[12], following which mitochondria were analysed by SDS–PAGE or BN–PAGE as described above.

Quantitative mass spectrometry using mitochondrial and whole-cell starting material, and data analysis. For NDUFV3, NDUFS6, NDUFA2, NDUFA8, NDUFA1, NDUFS5, NDUFC1, NDUFB4, NDUFB7, NDUFB10 and NDUFB11 knockouts, mass spectrometry was performed with SILAC-labelled whole-cell starting material as described previously[42] with modifications. In brief, cells cultured in 'heavy' $^{13}C_6{}^{15}N_4$-arginine, $^{13}C_6{}^{15}N_2$-lysine-containing or 'light' SILAC DMEM[15] were collected, washed in PBS and protein content determined by BCA assay. Measurements were performed in batches of 3–4 knockout cell lines in triplicate with a label switch. Each batch used a single pool of clonal HEK293T cells (1 sample grown in heavy DMEM, and 2 independent samples grown in

light DMEM) and knockout cell lines were grown with the complementary label orientation (1 in light DMEM, and 2 in heavy DMEM). Equal amounts of heavy and light (typically 250 μg) control HEK293T and knockout cells were mixed, and cells were solubilized in 1% (w/v) sodium deoxycholate, 100 mM Tris-HCl (pH 8.1). Lysates were sonicated for 30 min at 60 °C in a sonicator waterbath, followed by denaturation and alkylation through the addition of 5 mM Tris(2-carboxyethy)phosphine (TCEP), 20 mM chloroacetamide and incubation for 5 min at 99 °C with vortexing. Samples were digested with trypsin overnight at 37 °C. Detergent was removed by ethyl acetate extraction in the presence of 2% formic acid (FA), following which the aqueous phase was concentrated by vacuum centrifugation. Peptides were reconstituted in 0.5% FA and loaded onto pre-equilibrated small cation exchange (Empore Cation Exchange-SR, Supelco Analytical), stage-tips made in-house. Tips were washed with 6 load volumes of 20% acetonitrile (ACN), 0.5% FA and eluted in 5 sequential fractions of increasing amounts (45-300 mM) of ammonium acetate, 20% ACN, 0.5% FA. A sixth elution was collected using 5% ammonium hydroxide, 80% ACN following which fractions were concentrated, desalted and reconstituted as previously described[15].

Peptides were reconstituted in 0.1% trifluoroacetic acid (TFA) and 2% ACN and fractions analysed sequentially by online nano-HPLC/electrospray ionization-MS/MS on a Q Exactive Plus connected to an Ultimate 3000 HPLC (Thermo-Fisher Scientific). Peptides were first loaded onto a trap column (Acclaim C_{18} PepMap nano Trap × 2 cm, 100-μm I.D, 5-μm particle size and 300-Å pore size; ThermoFisher Scientific) at 15 μl min^{-1} for 3 min before switching the pre-column in line with the analytical column (Acclaim RSLC C_{18} PepMap Acclaim RSLC nanocolumn 75 μm × 50 cm, PepMap100 C_{18}, 3-μm particle size 100-Å pore size; ThermoFisher Scientific). The separation of peptides was performed at 250 nl min^{-1} using a nonlinear ACN gradient of buffer A (0.1% FA, 2% ACN) and buffer B (0.1% FA, 80% ACN), starting at 2.5% buffer B to 35.4% followed by ramp to 99% over 120 min (runs had a total acquisition time of 155 min to accommodate void and equilibration volumes). Data were collected in positive mode using Data Dependent Acquisition using m/z 375–1800 as MS scan range, HCD for MS/MS of the 12 most intense ions with $z \geq 2$. Other instrument parameters were: MS1 scan at 70,000 resolution (at 200 m/z), MS maximum injection time 50 ms, AGC target 3E6, Normalized collision energy was at 27% energy, Isolation window of 1.8 Da, MS/MS resolution 17,500, MS/MS AGC target of 1E5, MS/MS maximum injection time 100 ms, minimum intensity was set at 1E3 and dynamic exclusion was set to 15 s.

For the remaining knockouts, we used isolated mitochondria as starting material. Cells were cultured in SILAC DMEM as above and mitochondrial isolations performed in batches of 1–6 knockout cell lines in triplicate. Each batch contained a single set of clonal HEK293T mitochondria (2 independent isolations from heavy and 1 from light cells), with knockout mitochondria having the complementary label orientation (2 independent isolations from light DMEM and 1 from heavy cells). Mitochondria were isolated from cell pellets stored at −80 °C as previously described[43], but with modifications. Cells were resuspended in 20 mM HEPES-KOH (pH 7.6), 220 mM mannitol, 60 mM sucrose, 1 mM EDTA, 1 mM PMSF and homogenized as described above. The homogenate was centrifuged at 800g for 5 min at 4°C, and the supernatant again centrifuged at 10,000g for 10 min at 4 °C. Crude mitochondria were resuspended in the above buffer and the two differential centrifugation steps repeated. The resuspended pellet was then layered onto a sucrose cushion consisting of 10 mM HEPES-KOH (pH 7.6), 500 mM sucrose, 1 mM EDTA. Samples were centrifuged at 10,000g for 10 min at 4 °C, following which the protein concentration was estimated by BCA assay. Equal amounts of heavy and light (typically 20 μg) control HEK293T and knockout mitochondria were mixed as described above, collected by centrifugation at 18,000g and solubilized in 8 M urea, 50 mM ammonium bicarbonate. Proteins were acetone-precipitated, reduced and alkylated and desalted as previously described[15]. Peptides reconstituted in 0.1% TFA and 2% ACN were analysed on a Q Exactive Plus, or a LTQ-Orbitrap Elite Instrument. Instrument and method parameters for Q Exactive Plus were as described above, however, used a shorter gradient (90 min separation, 120 min total acquisition). For the Orbitrap Elite, instrument and method parameters were as previously described[15]. A single technical re-injection was collected for all mitochondrial samples.

All raw file names included identifiers for the batch, instrument and gradient used, knockout cell line being studied, and applicable label orientation. Raw files were analysed using the MaxQuant platform[44] version 1.5.4.1, searching against the Uniprot human database containing reviewed, canonical and isoform variants in FASTA format (June 2015) and a database containing common contaminants by the Andromeda search engine[45]. Default search parameters for an Arg10- and Lys8-labelled experiment were used with modifications. In brief, cysteine carbamidomethylation was used as a fixed modification, and N-terminal acetylation and methionine oxidation were used as variable modifications. False discovery rates of 1% for proteins and peptides were applied by searching a reverse database, and 're-quantify' and 'match from and to', 'match between runs' options were enabled with a match time window of 2 min. Experimental groups based on data gathered using different instrumentation and/or acquisition parameters were given odd numbered fractions to avoid falsely matched identifications, whereas fractionated whole-cell samples were given sequential fraction numbers. Unique and razor peptides with a minimum ratio count of 2 were used for quantification.

Using the Perseus platform (version 1.5.4.1), identifications were matched to the MitoCarta2.0 database[19] using Ensembl ENSG id and gene name identifiers. Identifications labelled by MaxQuant as 'only identified by site', 'reverse' and 'potential contaminant' were removed. Proteins having <3 valid values in a single experimental group were removed. For mitochondrial samples, we found the correlation of $\log_2$-ratio data from biological replicates in the same experimental group to be moderate at best and as low as 0.3 in some cases. We surmised the main cause of this to be batch and labelling effect, the former due to differences in mitochondrial isolations between batches and latter due to one (of three) replicates within each experimental group always being subjected to a label switch. To account for these and potentially other factors, we adopted an approach that borrows principles from RUV-2 (ref. 46) and SVA[47] methods for removing unwanted variations, with modifications in the algorithm for choosing the control proteins (that is, those not found in MitoCarta 2.0; ref. 19) and moderating the amount of adjustment for genes with small sample size due to missing values. Adjustments were performed in the R framework, following which the adjusted ratios were imported back into Perseus. The $\log_2$ ratio values for proteins in replicates were normally distributed and had equal variances. The mean $\log_2$-transformed ratios for each experimental group were calculated along with their standard deviation and P-value based on single sample two-sided t-test[15]. This statistical approach was consistent with published quantitative SILAC analyses employing similar instrumentation and methods[15,48,49]. Groups having <2 valid values were converted to 'NaN' (not a number). A quality matrix was generated based on the standard deviation, and corresponding values having a standard deviation greater than 1 converted to 'NaN'. This threshold was determined empirically to remove outliers from the main distribution of standard deviations across all samples. These data can be found in Supplementary Table 5.

Figures 3b and Extended Data Figs 6a, 8c and 9d were generated from a matrix containing $\log_2$-transformed median SILAC ratios having a standard deviation <1 for complex I subunits (Supplementary Table 7) and data were mapped to homologous subunits (Protein Data Bank accession 5LDW)[9]. For Fig. 3a, hierarchical clustering on rows (proteins) was performed using Pearson distance and average linkage. Data were pre-processed using k-means (clusters = 300). Images were generated using the PyMOL Molecular Graphics System, version 1.7.2.1 (Schrödinger, LLC). $\log_2$ SILAC ratios for some proteins in their corresponding knockout cell line had very low (generally >4-fold reduction) ratios, whereas others were reported NaN. This could be either due to the 're-quantify' option being turned on for the MaxQuant search, which results in translation of peak shapes from an identified isotope pattern being translated to its unidentified label partner, or indels in some lines generating a non-functional (but still translated) protein as we have seen previously[15].

For the identification of proteins dysregulated between knockouts of discrete modules (Fig. 4c, Supplementary Tables 8 and 9), triplicate $\log_2$-transformed SILAC ratios from Supplementary Table 5 were assigned to one of two groups based on the knockout being associated with the indicated module. Groups tested had comparable variance, and a modified Welch's two-sample t-test with permutation-based FDR statistics[50,51] was used to determine significance. Parameters for the test were: 70% minimum valid values, 250 permutations and significance being an FDR of <0.05.

For the Gene Ontology enrichment analysis in Fig. 2c, proteins with a $P < 0.05$ and with >1.5-fold change up or down were submitted to the DAVID online tool (https://david.abcc.ncifcrf.gov/home.jsp) for enriched biological processes (GOTERM_BP_FAT) and molecular function (GOTERM_MF_FAT). Functional annotation charts were exported and visualized using Cytoscape (version 3.4.0) and the Enrichment Map app[52] (version 2.1.0; $P < 0.005$). Contents of enriched terms indicated in Fig. 2c are detailed in Extended Data Fig. 5d.

Affinity-enrichment mass spectrometry and data analysis. Affinity-enrichment experiments in Fig. 4e, Extended Data Figs 3d and 4d and Supplementary Tables 2–4 and 10, 11 were performed from HEK293T and knockout cells complemented with the Flag-tagged protein cultured in heavy or light SILAC DMEM as previously described[15]. Mass spectrometry was performed on a Q Exactive Plus as above but using a shorter gradient (25 min separation, 60 min total acquisition). For data analysis, raw files were analysed using the MaxQuant platform as above using default search parameters for a Arg10 and Lys8 labelled experiment, with modifications. In brief, cysteine carbamidomethylation was used as a fixed modification, and N-terminal acetylation and methionine oxidation were used as variable modifications. False discovery rates of 1% for proteins and peptides

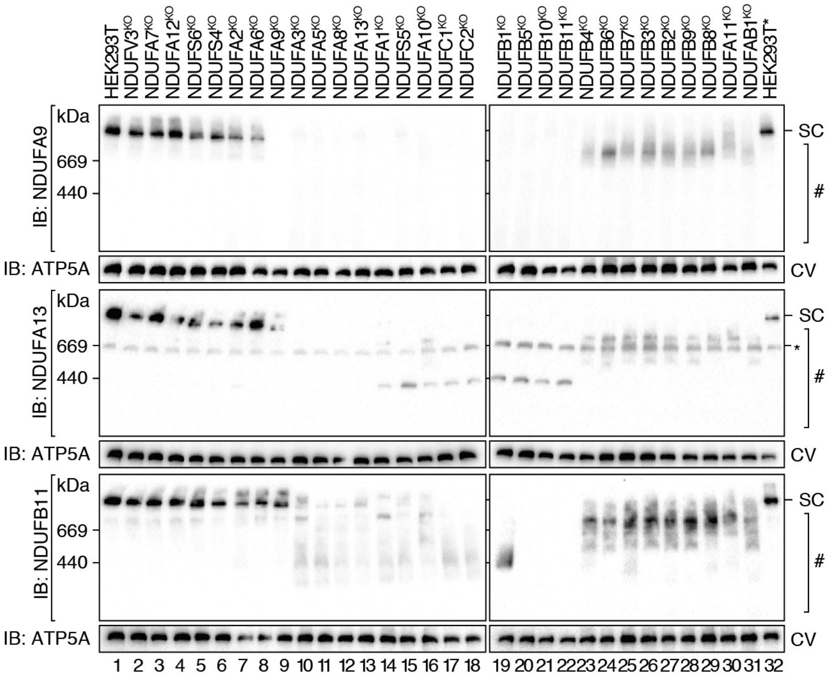
were applied by searching a reverse database, and 're-quantify' and 'match from and to', 'match between runs' options were enabled with a match time window of 2 min. Unique and razor peptides with a minimum ratio count of 1 were used for quantification. Data analysis was performed using the Perseus framework. Identifications were matched to MitoCarta2.0 data set[19] as above. Only proteins with a sequence coverage of 2 or more unique peptides were included in further analysis. Normalized SILAC ratios were inverted to achieve the orientation Flag-tagged/HEK293T and proteins not present in $>2/3$ replicates were removed. $\log_{10}$-transformed values had a normal distribution and comparable variance. For affinity-enrichment experiments, statistical method, sample size and analysis approaches were chosen based on published quantitative affinity-enrichment analyses employing similar instrumentation and methods[15,21,53,54]. P values were calculated by a single (Flag-tagged cell line enriched)-sided t-test and the negative logarithmic P-value plotted against the mean of the three replicates.

Miscellaneous molecular biology. cDNA inserts were obtained from an in-house cDNA library generated from our clonal HEK293T line. Briefly, RNA was isolated using TRIzol Reagent (ThermoFisher) according to manufacturer's instructions. The Superscript III first strand synthesis kit (ThermoFisher Scientific) was used to generate cDNA primed with either Oligo(dT) or random hexamers. Inserts were amplified from the library using Q5 High Fidelity DNA Polymerase (NEB) and Gibson assembled into the relevant plasmid (see above) using the NEBuilder HiFi DNA Assembly Master Mix (NEB) according to manufacturer's instructions. Sanger sequencing was performed from PCR product or plasmid template DNA. DNA sequence assembly and alignment to sequencing reads was performed using SnapGene (GSL Biotech) and Geneious (Biomatters).

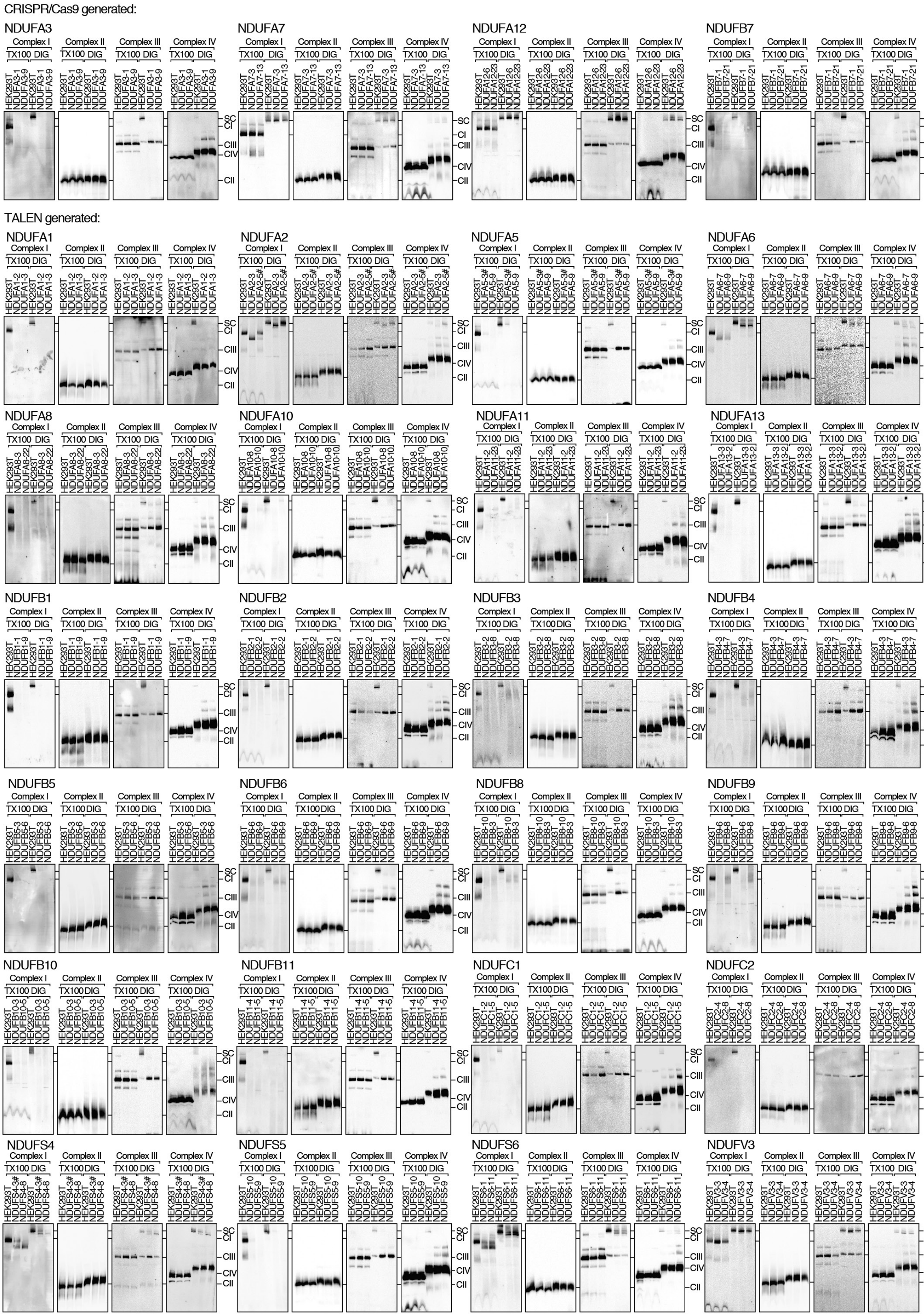
Immunofluorescence microscopy was performed as previously described[55] using primary antibodies (Flag or TOMM20) at 1:500 dilutions. Primary antibodies were labelled with anti-mouse conjugated Alexa Fluor 488 and anti-rabbit conjugated Alexa Fluor 568 secondary antibodies (Molecular Probes). Hoechst ($1\,\mu g\,ml^{-1}$) was used to stain nuclei. Cells were visualized using a Leica TCS SP8 equipped with HyD detectors. Images were processed using Image J[56]. All figures were prepared using Adobe Photoshop and Illustrator (CC2015.5).

Data reporting. No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

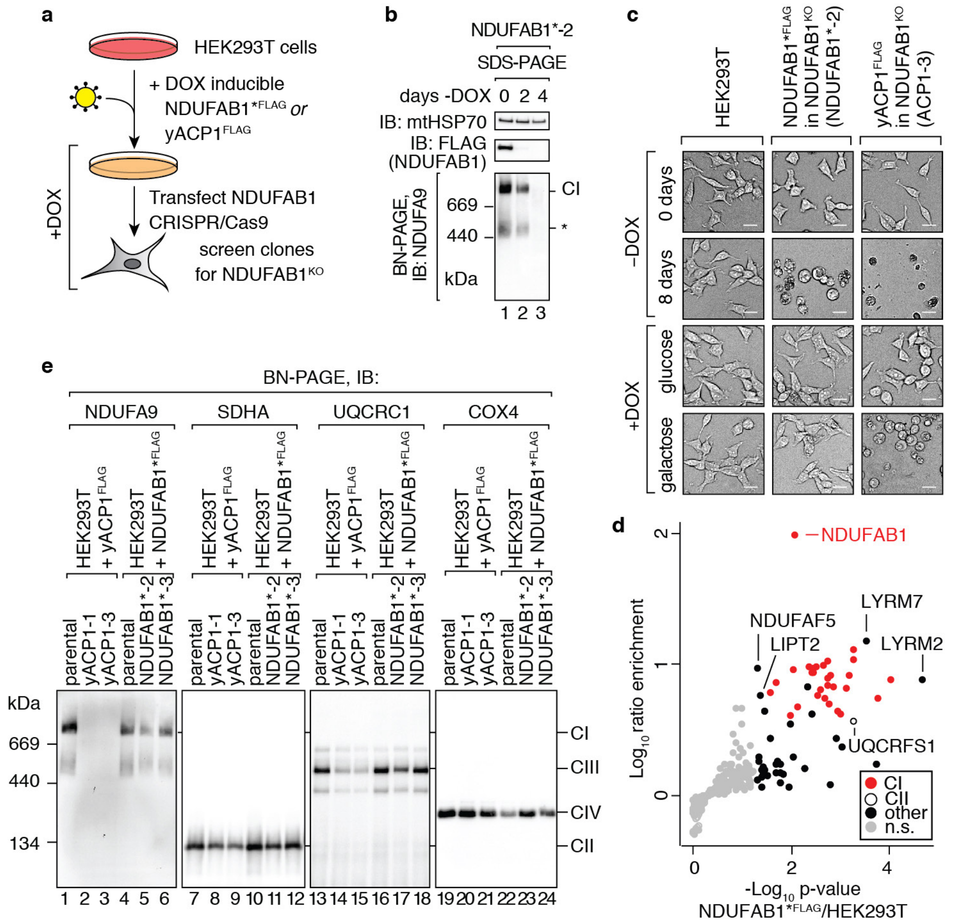
23. DuBridge, R. B. *et al.* Analysis of mutation in human cells by using an Epstein-Barr virus shuttle system. *Mol. Cell. Biol.* **7,** 379–387 (1987).
24. Heide, H. *et al.* Complexome profiling identifies TMEM126B as a component of the mitochondrial complex I assembly complex. *Cell Metab.* **16,** 538–549 (2012).
25. Vogel, R. O. *et al.* Identification of mitochondrial complex I assembly intermediates by tracing tagged NDUFS3 demonstrates the entry point of mitochondrial subunits. *J. Biol. Chem.* **282,** 7582–7590 (2007).
26. Huttlin, E. L. *et al.* The BioPlex Network: a systematic exploration of the human interactome. *Cell* **162,** 425–440 (2015).
27. Reyon, D. *et al.* FLASH assembly of TALENs for high-throughput genome editing. *Nat. Biotechnol.* **30,** 460–465 (2012).
28. Reljić, B. & Stroud, D. A. Screening strategies for TALEN-mediated gene disruption. *Methods Mol. Biol.* **1419,** 231–252 (2016).
29. Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat. Protocols* **8,** 2281–2308 (2013).
30. Sander, J. D. *et al.* ZiFiT (Zinc Finger Targeter): an updated zinc finger engineering tool. *Nucleic Acids Res.* **38,** W462–8 (2010).
31. Montague, T. G., Cruz, J. M., Gagnon, J. A., Church, G. M. & Valen, E. CHOPCHOP: a CRISPR/Cas9 and TALEN web tool for genome editing. *Nucleic Acids Res.* **42,** W401–7 (2014).
32. Morgenstern, J. P. & Land, H. Advanced mammalian gene transfer: high titre retroviral vectors with multiple drug selection markers and a complementary helper-free packaging cell line. *Nucleic Acids Res.* **18,** 3587–3596 (1990).
33. Acín-Pérez, R., Fernández-Silva, P., Peleato, M. L., Pérez-Martos, A. & Enriquez, J. A. Respiratory active mitochondrial supercomplexes. *Mol. Cell* **32,** 529–539 (2008).
34. McKenzie, M., Lazarou, M., Thorburn, D. R. & Ryan, M. T. Analysis of mitochondrial subunit assembly into respiratory chain complexes using Blue Native polyacrylamide gel electrophoresis. *Anal. Biochem.* **364,** 128–137 (2007).
35. Wittig, I., Braun, H. P. & Schägger, H. Blue native PAGE. *Nat. Protocols* **1,** 418–428 (2006).
36. Schägger, H. & von Jagow, G. Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. *Anal. Biochem.* **166,** 368–379 (1987).
37. Ryan, M. T., Voos, W. & Pfanner, N. Assaying protein import into mitochondria. *Methods Cell Biol.* **65,** 189–215 (2001).
38. Dunning, C. J. *et al.* Human CIA30 is involved in the early assembly of mitochondrial complex I and mutations in its gene cause disease. *EMBO J.* **26,** 3227–3237 (2007).
39. Jänicke, A., Vancuylenberg, J., Boag, P. R., Traven, A. & Beilharz, T. H. ePAT: a simple method to tag adenylated RNA to measure poly(A)-tail length and other 3′ RACE applications. *RNA* **18,** 1289–1295 (2012).
40. Harrison, P. F. *et al.* PAT-seq: a method to study the integration of 3′-UTR dynamics with gene expression in the eukaryotic transcriptome. *RNA* **21,** 1502–1510 (2015).
41. Frazier, A. E. & Thorburn, D. R. Biochemical analyses of the electron transport chain complexes by spectrophotometry. *Methods Mol. Biol.* **837,** 49–62 (2012).
42. Kulak, N. A., Pichler, G., Paron, I., Nagaraj, N. & Mann, M. Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. *Nat. Methods* **11,** 319–324 (2014).
43. Johnston, A. J. *et al.* Insertion and assembly of human Tom7 into the preprotein translocase complex of the outer mitochondrial membrane. *J. Biol. Chem.* **277,** 42197–42204 (2002).
44. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26,** 1367–1372 (2008).
45. Cox, J. *et al.* Andromeda: a peptide search engine integrated into the MaxQuant environment. *J. Proteome Res.* **10,** 1794–1805 (2011).
46. Gagnon-Bartsch, J. A. & Speed, T. P. Using control genes to correct for unwanted variation in microarray data. *Biostatistics* **13,** 539–552 (2012).
47. Leek, J. T. & Storey, J. D. Capturing heterogeneity in gene expression studies by surrogate variable analysis. *PLoS Genet.* **3,** 1724–1735 (2007).
48. Münch, C. & Harper, J. W. Mitochondrial unfolded protein response controls matrix pre-RNA processing and translation. *Nature* **534,** 710–713 (2016).
49. Wrobel, L. *et al.* Mistargeted mitochondrial proteins activate a proteostatic response in the cytosol. *Nature* **524,** 485–488 (2015).
50. Hubner, N. C. *et al.* Quantitative proteomics combined with BAC TransgeneOmics reveals in vivo protein interactions. *J. Cell Biol.* **189,** 739–754 (2010).
51. Tusher, V. G., Tibshirani, R. & Chu, G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl Acad. Sci. USA* **98,** 5116–5121 (2001).
52. Merico, D., Isserlin, R., Stueker, O., Emili, A. & Bader, G. D. Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. *PLoS One* **5,** e13984 (2010).
53. Stroud, D. A. *et al.* Composition and topology of the endoplasmic reticulum-mitochondria encounter structure. *J. Mol. Biol.* **413,** 743–750 (2011).
54. Gebert, N. *et al.* Dual function of Sdh3 in the respiratory chain and TIM22 protein translocase of the mitochondrial inner membrane. *Mol. Cell* **44,** 811–818 (2011).
55. Richter, V. *et al.* Structural and functional analysis of MiD51, a dynamin receptor required for mitochondrial fission. *J. Cell Biol.* **204,** 477–486 (2014).
56. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9,** 671–675 (2012).
57. Ostergaard, E. *et al.* Respiratory chain complex I deficiency due to NDUFA12 mutations as a new cause of Leigh syndrome. *J. Med. Genet.* **48,** 737–740 (2011).
58. Assouline, Z. *et al.* A constant and similar assembly defect of mitochondrial respiratory chain complex I allows rapid identification of NDUFS4 mutations in patients with Leigh syndrome. *Biochim. Biophys. Acta* **1822,** 1062–1069 (2012).
59. Haack, T. B. *et al.* Mutation screening of 75 candidate genes in 152 complex I deficiency cases identifies pathogenic variants in 16 genes including NDUFB9. *J. Med. Genet.* **49,** 83–89 (2012).
60. Kirby, D. M. *et al.* NDUFS6 mutations are a novel cause of lethal neonatal mitochondrial complex I deficiency. *J. Clin. Invest.* **114,** 837–845 (2004).
61. van den Bosch, B. J. *et al.* Defective NDUFA9 as a novel cause of neonatally fatal complex I disease. *J. Med. Genet.* **49,** 10–15 (2012).
62. Hoefs, S. J. *et al.* NDUFA10 mutations cause complex I deficiency in a patient with Leigh disease. *Eur. J. Hum. Genet.* **19,** 270–274 (2011).
63. Angebault, C. *et al.* Mutation in NDUFA13/GRIM19 leads to early onset hypotonia, dyskinesia and sensorial deficiencies, and mitochondrial complex I instability. *Hum. Mol. Genet.* **24,** 3948–3955 (2015).
64. Haack, T. B. *et al.* Molecular diagnosis in mitochondrial complex I deficiency using exome sequencing. *J. Med. Genet.* **49,** 277–283 (2012).
65. Shehata, B. M. *et al.* Exome sequencing of patients with histiocytoid cardiomyopathy reveals a de novo *NDUFB11* mutation that plays a role in the pathogenesis of histiocytoid cardiomyopathy. *Am. J. Med. Genet. A.* **167A,** 2114–2121 (2015).
66. Fernandez-Moreira, D. *et al.* X-linked *NDUFA1* gene mutations associated with mitochondrial encephalomyopathy. *Ann. Neurol.* **61,** 73–83 (2007).
67. Peralta, S. *et al.* Partial complex I deficiency due to the CNS conditional ablation of Ndufa5 results in a mild chronic encephalopathy but no increase in oxidative damage. *Hum. Mol. Genet.* **23,** 1399–1412 (2014).



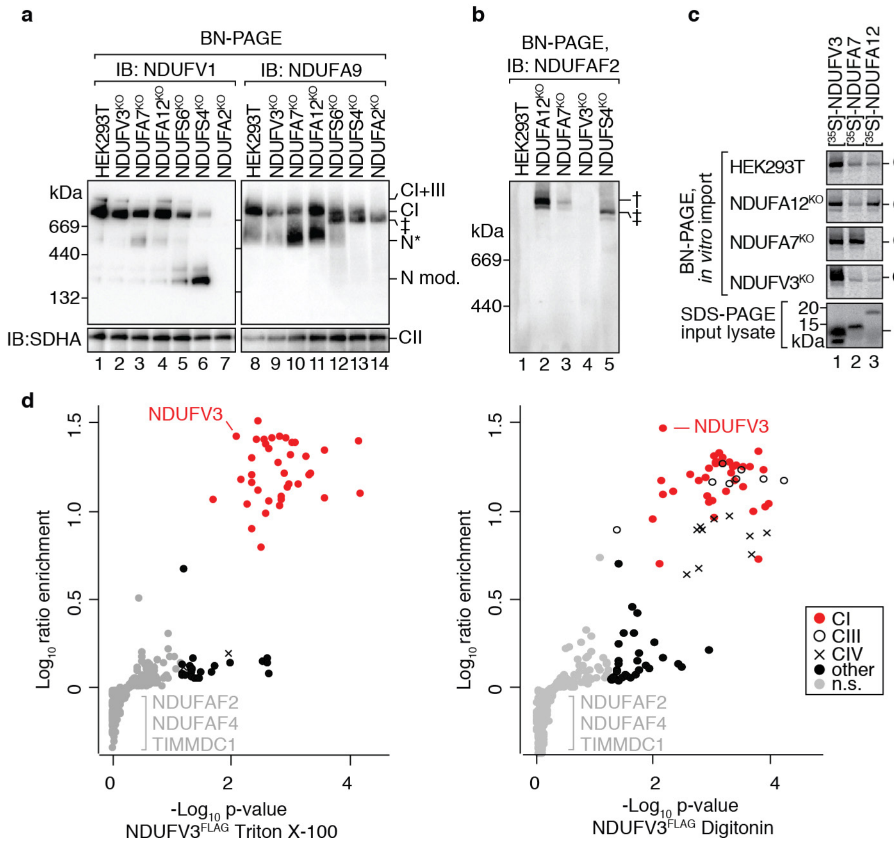
Extended Data Figure 1 | Assembly analysis of the complex I/III/IV supercomplex in knockout cell lines. Mitochondria were solubilized in digitonin and complexes separated by BN–PAGE followed by immunoblotting using the indicated antibodies. An antibody against complex V (CV) subunit ATP5A was used as loading control. #, subcomplexes; *, non-specific. SC, supercomplex.



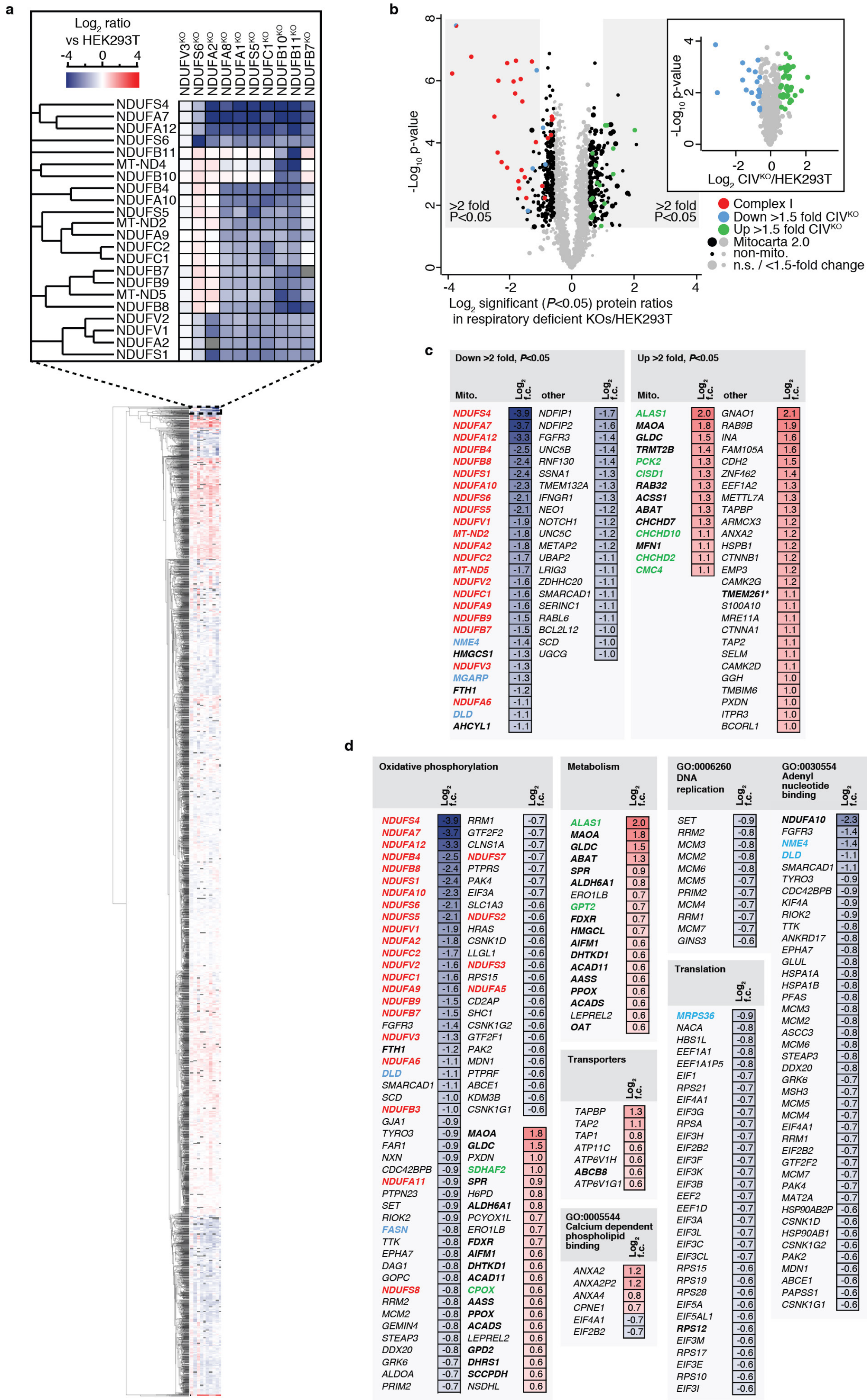
Extended Data Figure 2 | Steady-state levels of respiratory chain complexes I–IV and supercomplex forms in the 28 complex I accessory subunit knockout lines generated in this study. NDUFA9^{KO} has been analysed previously[20], whereas the NDUFAB1^{KO} is described in Extended Data Fig. 3. Mitochondria were solubilized in Triton X-100 (TX100) or digitonin (DIG) and analysed by BN–PAGE and immunoblotting with antibodies against NDUFA9 (complex I), SDHA (complex II), UQCRC1 (complex III) and COX4 (complex IV). In Triton X-100 samples, some complex III–IV supercomplex is retained. #, secondary clone later identified be an incomplete knockout.



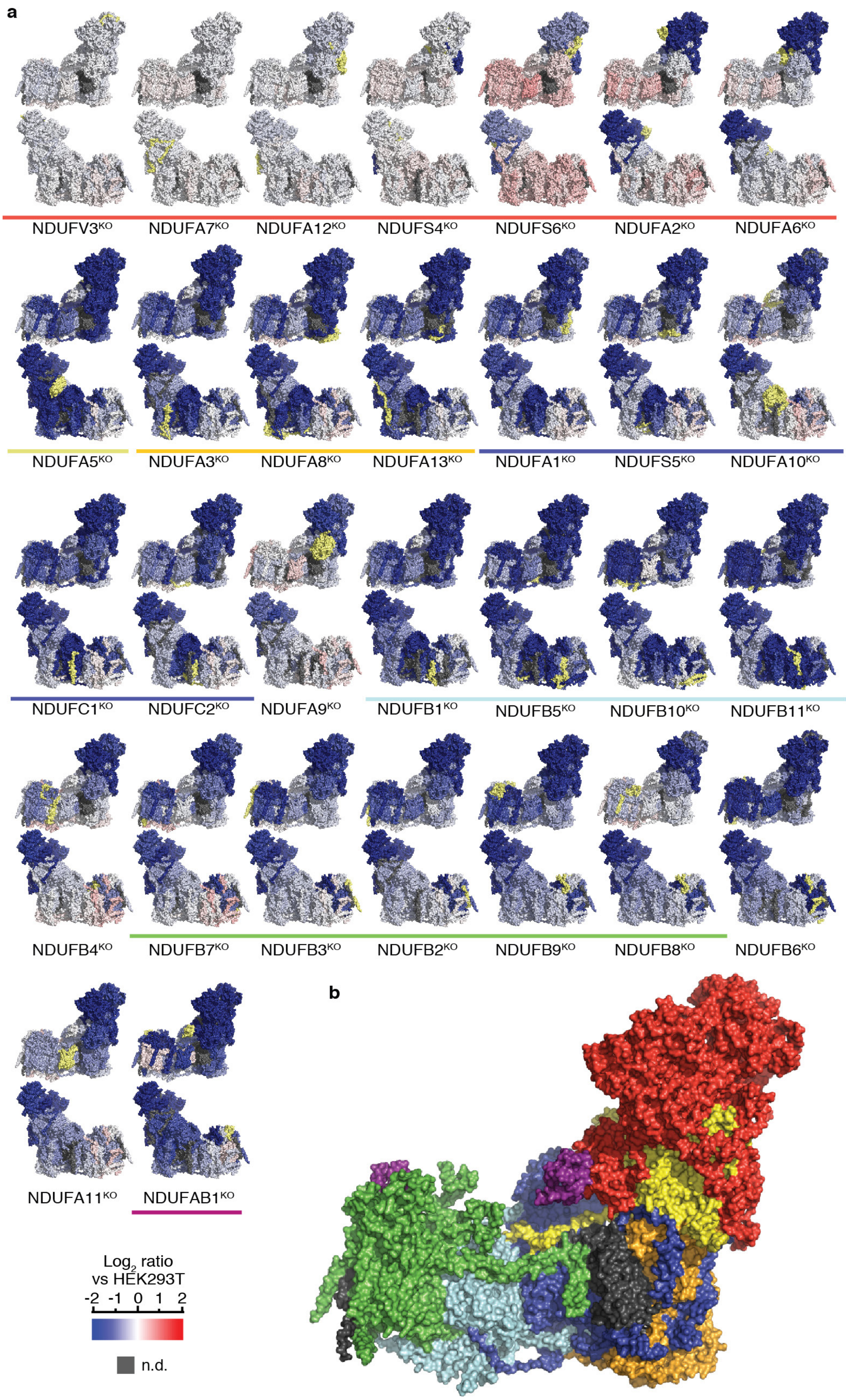
Extended Data Figure 3 | Generation and analysis of NDUFAB1-knockout cell lines. a, Scheme detailing knockout strategy of genomic *NDUFAB1* using doxycycline (DOX)-inducible expression of CRISPR/Cas9-resistant NDUFAB1 (NDUFAB1*Flag) or yACP1Flag. **b**, *NDUFAB1* knockouts complemented with NDUFAB1Flag (NDUFAB1*-2) cells were cultured in media lacking DOX for the indicated times. Isolated mitochondria were analysed by BN–PAGE (Triton X-100) or SDS–PAGE and immunoblotting with the indicated antibodies. **c**, Brightfield images of cells grown ± DOX, or +DOX in glucose or galactose cell culture medium. Scale bars, 25 μm. Representative results from 4 independent experiments. **d**, SILAC-labelled mitochondria from DOX-treated HEK293T or NDUFAB1Flag (NDUFAB1*-2) cells were solubilized in Triton X-100 and incubated with anti-Flag affinity gel. Elutions were mixed and analysed by liquid chromatography–mass spectrometry (LC–MS). Proteins enriched with NDUFAB1 include complex I subunits and LYRM proteins. *P* values are from an unpaired single-sided *t*-test. $n = 3$ biological replicates; light grey dots, not significant ($P > 0.05$). **e**, Mitochondria isolated from *NDUFAB1* knockouts complemented with yACP1Flag or NDUFAB1Flag were solubilized in Triton X-100 and analysed by BN–PAGE and immunoblotting with the indicated antibodies.



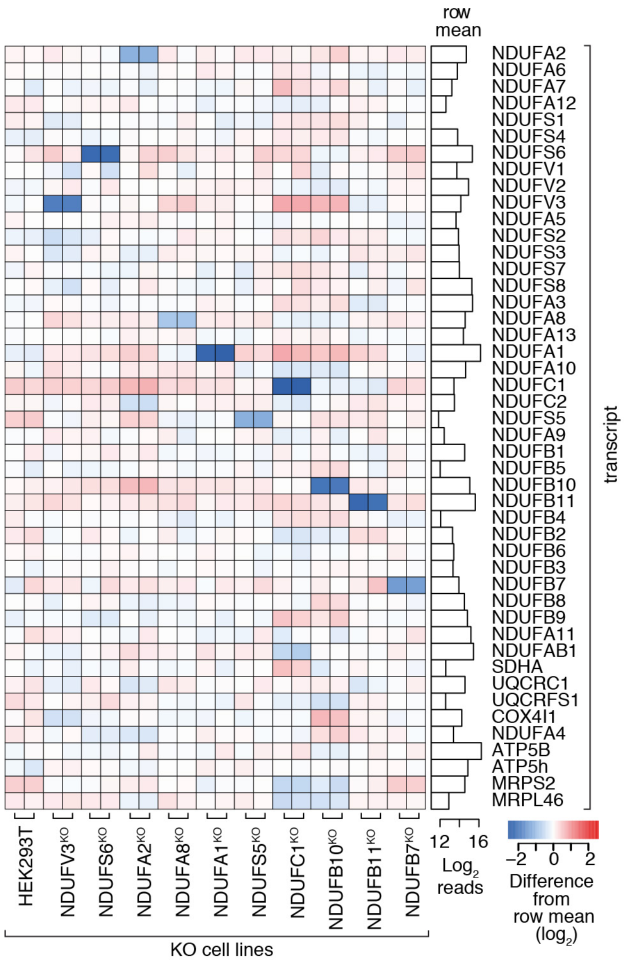
Extended Data Figure 4 | Analysis of N-module accessory subunits. **a**, Mitochondria were isolated from cell lines, solubilized in Triton X-100 and analysed by BN–PAGE and immunoblotting for N-module subunit NDUFV1 or non N-module subunit NDUFA9. ‡, complex lacking N-module; N*, subcomplex containing N-module. SDHA was used as a loading control. **b**, Mitochondria were solubilized in digitonin and analysed by BN–PAGE and immunoblotting for NDUFAF2. †, NDUFAF2 associated complex I. **c**, [^{35}S]methionine-labelled proteins were imported into the indicated mitochondria, solubilized in digitonin and analysed by BN–PAGE and autoradiography. 10% of the input lysate was analysed by SDS–PAGE and autoradiography. CI^{SC}, complex I supercomplex; *, non-specific band. **d**, Mitochondria isolated from NDUFV3-knockout cells complemented with NDUFV3^{Flag} were solubilized in Triton X-100 or digitonin and complexes bound to anti-Flag affinity gel. Eluted proteins were analysed by LC–MS. *P* values are from an unpaired single-sided *t*-test. $n = 3$ biological replicates; light grey dots, not significant ($P > 0.05$).



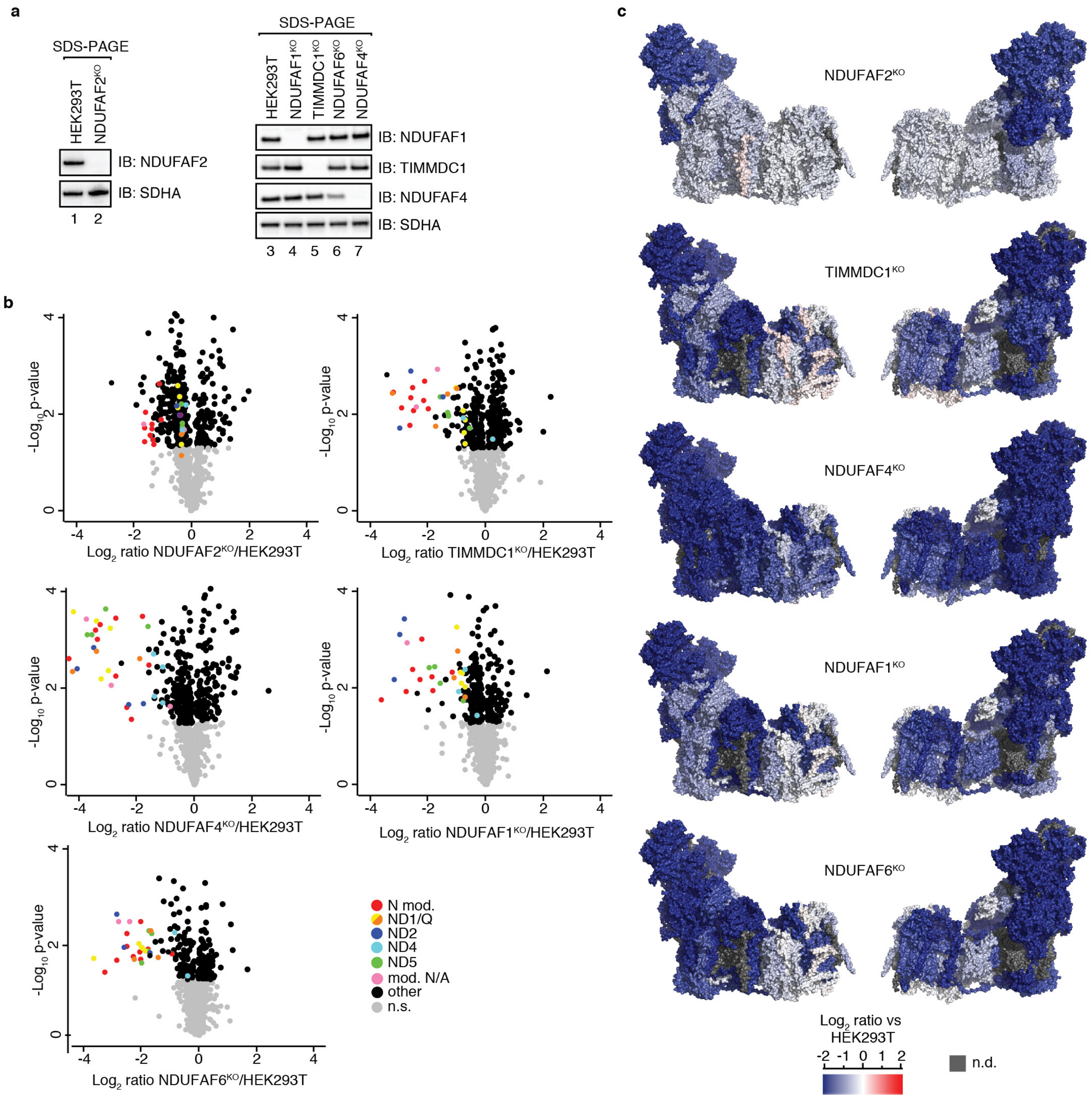
Extended Data Figure 5 | Proteomic analysis of knockout cell lines. **a**, Relative levels of proteins in representative accessory subunit knockout cell lines, clustered according to Euclidean distance. Column order is as in Fig. 2b. The inset shows complex I subunit-specific clusters. **b**, Volcano plot depicting proteins regulated in representative accessory subunit knockout cell lines containing respiration defects (NDUFA2, NDUFA8, NDUFS5, NDUFC1, NDUFB10, NDUFB11 and NDUFB7 knockouts). Proteins found to be regulated in a cell line with a severe complex IV defect[15] are shaded light blue (down) and green (up), suggesting their response is due to general defects in respiration. Inset, volcano plot depicting the relative level of proteins in a complex IV knockout cell line. *P* values are from an unpaired *t*-test; $n = 8$ independent means comprised each of 3 biological replicates (main panel), $n = 3$ (inset) biological replicates; light grey dots, not significant ($P > 0.05$, <1.5-fold change). Data are reproduced in Supplementary Table 6. **c**, Proteins affected >2-fold in levels in respiration-deficient subunit knockout cell lines. Colour key according to **b**. Bold, proteins listed in MitoCarta2.0. **d**, Proteins associated with GO terms and groups outlined in Fig. 2d.



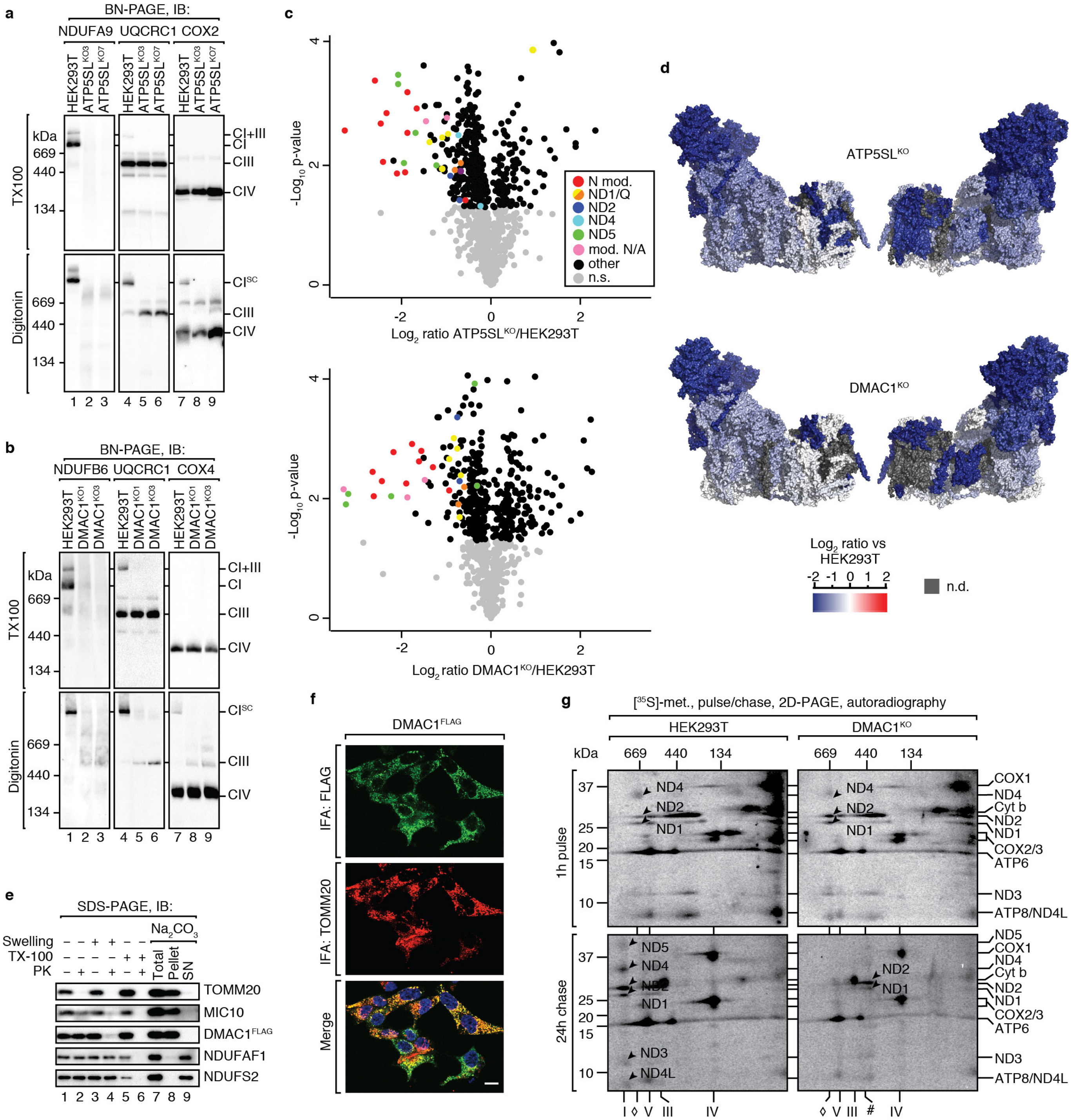
Extended Data Figure 6 | Mapping of complex I subunit levels onto the structure. **a**, Subunit levels in complex I accessory subunit knockout lines were mapped to homologous subunits in the bovine single-particle electron cryo-microscopy structure of complex I (ref. 9) as in Fig. 3b. Both sides of complex I are shown. Median ratio data used in the preparation of this figure can be found in Supplementary Table 7. **b**, Opposite side view of Fig. 3c. n.d., dark grey shading on the structures, subunits not quantified. Subunits not clustered to modules removed for clarity.



Extended Data Figure 7 | mRNA expression levels in selected accessory subunit knockout lines. Transcripts were measured for nuclear-encoded complex I subunit genes along with control genes from complex II (SDHA), complex III (UQCRC1, UQCRFS1), complex IV (COX4L1, NDUFA4), complex V (ATP5B, ATP5H) and mt-ribosome (MRPS2, MRPL46) in knockout lines (performed in duplicate).



Extended Data Figure 8 | Analysis of assembly factor knockout lines. **a**, Mitochondrial proteins from the indicated cell lines were separated by SDS–PAGE and subjected to western blot analysis. **b**, Volcano plots showing fold changes versus *P* values for the mitochondrial proteins in assembly factor knockout cell lines. *P* values are from an unpaired *t*-test; $n = 3$ biological replicates; coloured dots are according to the key at bottom right. n.s., not significant ($P > 0.05$). **c**, Subunit levels mapped to homologous subunits in the bovine single-particle electron cryo-microscopy structure as in Fig. 3b. n.d., dark grey shading on the structures, subunits not quantified. Both sides of complex I are shown.



Extended Data Figure 9 | Characterization of DMAC1 and ATP5SL. **a**, ATP5SL-knockout mitochondria were solubilized in Triton X-100 or digitonin and analysed by BN–PAGE and immunoblotting with the indicated antibodies. **b**, As in **a** using DMAC1-knockout mitochondria. **c**, Volcano plots showing fold changes versus *P* values for the mitochondrial proteins in ATP5SL and DMAC1 knockout cell lines. *P* values are from an unpaired *t*-test; $n = 3$ biological replicates; coloured dots represent complex I subunits depicted in the key; n.s., $P > 0.05$. **d**, Subunit levels mapped to homologous subunits in the bovine single-particle electron cryo-microscopy structure as for Fig. 3b. n.d., dark grey shading on the structures, not quantified. Both sides of complex I are shown. **e**, Mitochondria isolated from DMAC1 cells complemented with DMAC1^{Flag} were resuspended in isotonic buffer, hypoosmotic swelling buffer, or Triton X-100 followed by proteinase K (PK) incubation where indicated. Alternately, mitochondria were treated with 100 mM Na_2CO_3 and membrane-integral (pellet) and soluble or peripherally attached (supernatant, SN) proteins were separated by ultracentrifugation. Samples were analysed by SDS–PAGE and immunoblotting for TOMM20 (outer mitochondrial membrane protein); MIC10 (integral inner membrane protein exposed to intermembrane space); NDUFAF1 (matrix, soluble); and NDUFS2 (matrix, peripheral). **f**, DMAC1-knockout cells complemented with DMAC1^{Flag} were analysed by immunofluorescence microscopy with the indicated antibodies. Scale bar, 20 μm. Representative result from 3 independent experiments. **g**, Cells were pulsed with [^{35}S]methionine for 1 h and chased for the indicated times. Isolated mitochondria were solubilized in Triton X-100 and analysed by 2D-PAGE and autoradiography. ◊, 600 kDa complex; #, subcomplex containing ND1 and ND2.

Extended Data Table 1 | Pathogenic mutations in complex I accessory subunit genes in patients with mitochondrial disease

Gene Symbol	OMIM#	Mutations	Known/Predicted Impact on Subunit	Impact on Complex I Assembly
Mild Assembly Defects[a]				
NDUFA12	*614530	homozygous p.R60X	No detectable protein	Decreased CI assembly[57]
NDUFS4	*602694	homozygous p.K158fs homozygous p.W96X homozygous p.R106X homozygous p.W15X homozygous IVS1AS, G>A, -1 homozygous p.K154Kfs (and other similar homozygous or compound heterozygous mutations likely to be null mutations)	Predicted Null mutation Predicted Null mutation Predicted Null mutation Predicted Null mutation Predicted Null mutation Predicted Null mutation	All mutations studied result in partially assembled CI lacking N-module[58]
NDUFS6	*603848	homozygous IVS2DS, T-A, +2 homozygous 4.175-KB DEL, EX3-4DEL homozygous p.C115Y homozygous p.Q118X[59]	Predicted Null mutation Predicted Null mutation Uncertain Predicted Null mutation	All mutations studied result in partially assembled CI lacking N-module[60]
Severe Assembly Defects				
NDUFA9	*603834	Homozygous p.R321P	Marked decrease in NDUFA9 protein amount	Decreased CI assembly[61]
NDUFA10	*603835	homozygous p.G99E compound heterozygous p.Met1?/p.Q142R[62]	Uncertain ~10% of normal levels of NDUFA13 protein	Decreased CI assembly[62]
NDUFA11	*612638	homozygous IVS1DS, G-A, +5	“Leaky” - 2:1 ratio of wildtype to normal transcript	Not assessed
NDUFA13	*609435	homozygous p.R57H[63]	30-40% of normal levels of NDUFA13 protein	Decreased CI assembly[63]
NDUFB3	*603839	homozygous p.W22R compound heterozygous p.W22R/p.G70X	Uncertain 1 predicted Null mutation & 1 uncertain	Decreased CI assembly[64]
NDUFB9	*601445	homozygous p.L64P	Some residual protein	Not assessed
NDUFB11	*300403	de novo heterozygous p.R88X (female) heterozygous p.R134SfsX3 (female) de novo heterozygous p.Y108X (female)[65] de novo heterozygous p.W85X (female)[65]	X-linked gene: mixture of null and wildtype cells X-linked gene: mixture of null and wildtype cells X-linked gene: mixture of null and wildtype cells X-linked gene: mixture of null and wildtype cells	Not assessed
NDUFA1	*300078	hemizygous p.G8R (male) hemizygous p.R37S (male) hemizygous p.G32R (male) de novo heterozygous p.G32R (female)	X-linked gene: Uncertain X-linked gene: Uncertain X-linked gene: Uncertain X-linked gene: Uncertain	Decreased CI assembly[66]
NDUFA5	*601677	No patients reported but Ndufa5 knockout mice die around embryonic day-9		Decreased CI assembly[67]

[a]Three accessory subunits in which knockouts cause mild complex I assembly defects have had patients reported with pathogenic mutations; in almost all cases the mutations are expected to cause two null alleles, suggesting that almost complete loss of function of these subunits may be required to cause human disease. Eight accessory subunits in which knockouts cause severe complex I assembly defects have had patients reported with pathogenic mutations; in almost all cases the patients have at least one missense mutation or some evidence that some residual subunit protein is present. This suggests that complete loss of function of these subunits may not be compatible with human life. Interpretation of the data for the NDUFB11 and NDUFA1 subunits is complicated by their being encoded on the X chromosome. Males thus have only one copy of these genes whereas females have 2 copies, with some cells expressing the wild-type and some expressing the mutant allele. All reported NDUFB11 patients are female and had stop codon or frameshift mutations expected to cause null alleles. Such patients often have skewed X-chromosome inactivation, with most cells expressing the wild-type allele, which may compensate partly for the severity of the defect. An additional subunit, NDUFA5, has not had patients with mutations identified but knockout of the mutation in mice results in embryonic lethality. This is consistent with the suggestion that human fetuses may not be viable if they have null-type mutations in both alleles of genes encoding accessory subunits linked to severe assembly defects.
#Online Mendelian Inheritance in Man (OMIM). McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, MD) accessed 20 January 2016 (http://omim.org).
References 57–67 are cited in the table.

TOOLBOX

DEMOCRATIC DATABASES: SCIENCE ON GITHUB

Scientists are turning to a software-development site to share data and code.

ILLUSTRATION BY THE PROJECT TWINS

BY JEFFREY PERKEL

When the Ebola outbreak in West Africa picked up pace in July 2014, Caitlin Rivers started to collect data on the people affected. Rivers, then a PhD student in computational epidemiology, wanted to model the outbreak's spread. So every day she downloaded PDF updates released by the ministries of health of the virus-stricken countries, and converted the numbers into computer-readable tables. Rather than keeping these files to herself, she posted them to GitHub.com, a hugely popular website for collaborative work on software code. Rivers thought the postings might attract those interested in up-to-date information from the Ebola outbreak. "I figured if I needed it, other people would, too," she says.

Rivers was right. Other researchers began to download the data and contribute to the project. On some days, third parties would download and convert the ministries' data before her, and load them into the GitHub repository. Others created programming scripts to do simple error-checks on the data, such as ensuring that the daily patient counts made sense. At the time, GitHub was "really the only place on the Internet that you could interact with these data as data, and not as a PDF", says Rivers, who was at Virginia Polytechnic Institute and State University in Blacksburg when she began the project, and is now an epidemiologist at the US Army Public Health Center in Edgewood, Maryland.

Launched in 2008 to assist software developers, GitHub now boasts some 15 million users and is an increasingly popular site for researchers to share, maintain and update scientific data sets and code (see 'Growing influence of GitHub'). GitHub is "the biggest revelation in my workflow … since I started writing code", says Daniel Falster, a postdoctoral researcher in ecology at Macquarie University in Sydney, Australia. "When we started using GitHub, it was just amazing. We now use it in everything that we do." Falster's Biomass and Allometry Database, which aggregates various measures of plant size from 176 studies, is stored on the site. So is the Open Tree of Life project, which aims to compile different published phylogenies to build one master 'tree of life'. It uses GitHub to store data files and publication records, and to accept new data sets from third parties.

Plenty of websites are dedicated to sharing data. But GitHub is specifically designed for transparent, open collaboration because it ▶

uses version-control software to track every change made to code or data. This means that large, distributed teams of programmers can work together on a project online, and users can scroll back in time through a file's version history, seeing each change, when it was made, by whom and for what purpose. Programmers can copy ('fork') a repository to experiment with new ideas; useful changes can be folded into the main project, while others can be ignored or rolled back later.

For instance, anyone can visit the GitHub-based Open Exoplanet Catalogue — a growing database of the thousands of known planets outside the Solar System — and submit new information through their browser. As with the Open Tree of Life, the project's main website doesn't have github.com in the URL address, so casual visitors wouldn't necessarily know that they are interacting with version-control software — but the files are openly available in a GitHub repository for more sophisticated users. Making an edit alerts the project's developers, including Hanno Rein, an astrophysicist at the University of Toronto in Canada, to review the suggested change. GitHub, says Rein, allows for a "way more democratic system" than would a static online catalogue of exoplanets, because any user can suggest changes and can even customize a version of the data set to their own specifications. Some 100 people have forked the project's repository, and Rein's smartphone app Exoplanet, which runs off the same database, has attracted around 10 million downloads.

FROM LINUX TO THE LAB

The software tool that GitHub relies on is called Git. It was created in 2005 by coder Linus Torvalds to manage development of the open-source operating system Linux — a huge project that involved thousands of independent programmers. "Git is a technology that's designed for very fine-grained, line-by-line monitoring of changes in source code," says Arfon Smith, a program manager for GitHub in Seattle, Washington. It is not the only version-control software available (another option is Mercurial), but it is one of the most popular.

Many programmers use Git on their own computers. For scientist coders, the tool works like a laboratory notebook for scientific computing, says Katy Huff, a nuclear engineer at the University of Illinois at Urbana–Champaign: just like a lab notebook, it keeps a lasting record of events. But its syntax and workflow are notoriously confusing. "I'm comfortable saying that the interface is unnecessarily non-intuitive," Huff says.

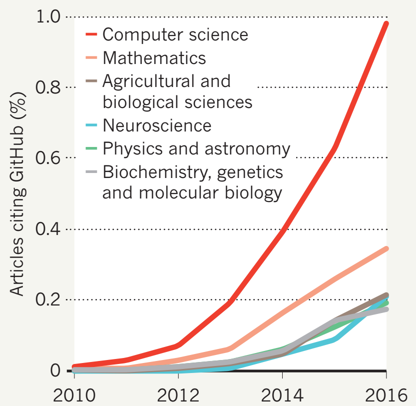
GitHub's prettier browser interface softens some of Git's hard edges, making it easier for novices to contribute. The site now hosts millions of projects, some personal, some massively collaborative, and is free for open-source projects. (Users and organizations that want to keep their files private pay US$7 per month and up. A related service called Bitbucket, which also runs on Git, offers unlimited free public and private repositories for up to five users; larger collaborations cost from $10 per month.)

Not every kind of data set works well with Git software. The tool records line by line how files have changed. It works well with text files such as source code, XML files, manuscripts written in Markdown or LaTeX, and CSV files (which can be exported from Excel, for instance). But it cannot effectively keep track of changes in non-human-readable 'binary' files, such as Microsoft Office documents and images, because the program's 'diff'ing' function, which identifies how files change from version to version, cannot interpret such data. "As soon as you introduce a binary format that isn't line-oriented, Git does a terrible, terrible job of versioning that content," Smith says.

GitHub also imposes file limitations; it has a hard limit of 100 megabytes per file, and a 'soft' cap of a gigabyte per repository. (A plug-in called Large File Storage allows Git and GitHub to more effectively handle larger files, although it still cannot report the differences between binary versions.)

GROWING INFLUENCE OF GITHUB

An increasing proportion of research articles cite GitHub in their references.



RICHARD VAN NOORDEN/SOURCE: ELSEVIER SCOPUS DATABASE

FAST AND FLEXIBLE

GitHub makes most sense for those researchers working with relatively small, text-based data sets that are being actively updated, curated and maintained by groups of scientists — such as Rivers' Ebola-virus project. Nick Loman, a microbial genomicist and bioinformatician at the University of Birmingham, UK, has also used the site to drive fast-paced studies of pathogens. Loman is a member of the ZiBRA (Zika in Brazil Real Time Analysis) project, an ongoing Brazilian surveillance effort that collects Zika-virus samples across the country and sequences and analyses them in real time. Traditionally, Loman says, DNA sequence data go to archives such as GenBank — and these data will too. But it can take time for those sites to release data to the public. GitHub, he says, provided a faster and more flexible way to disseminate draft data sets, rather like tweeting a research finding in advance of publication.

Because data sets on GitHub can be changed or deleted by their authors, the site doesn't guarantee a permanently citable archive, warns Smith. Those interested in creating a long-term, permanent record of their data set as it exists at a particular point in time — for example, when a paper is published — should consider storing the relevant version of their data on dedicated scientific sites, such as Zenodo and Figshare. Both of these sites allow GitHub users to archive snapshots of their repositories, and will provide a citable Digital Object Identifier (DOI) for the data set. According to Smith, some 8,000 GitHub users have done so.

Another data-sharing option is Dat, a general-purpose tool for sharing and syncing data between different computers. According to lead programmer Max Ogden in Portland, Oregon, Dat provides versioning in a similar way to Git for collaborative work, but includes a peer-to-peer file-sharing system for distributing data files. Ogden says that Dat is more adept at handling large binary files because it breaks them into chunks and transfers only those pieces that have changed.

Data sharing is a key requirement of open science, and researchers can share data sets anywhere they wish. But even if they don't use GitHub.com, scientists should consider using Git or a comparable tool to record changes to data sets and data-processing scripts, says Tracy Teal, executive director of Data Carpentry, a non-profit organization that trains researchers in working with data. Researchers interested in learning to use Git and GitHub have many online resources to turn to: Codecademy offers a free interactive tutorial, as does GitHub (try.github.io). Greg Wilson, founder of the research-computing skills site Software Carpentry, co-authored a how-to guide in January (J. D. Blischak *et al. PLoS Comput. Biol.* **12,** e1004668; 2016). And many programmers and bioinformaticians use Git — so they, too, can always be asked for help.

Despite their steep learning curves, Git and GitHub have a loyal fan base among scientists. Emily Jane McTavish, an evolutionary biologist at the University of California, Merced, and a member of the Open Tree of Life project, says it's an essential resource. "I don't know how I lived without it." ■

CORRECTION

The article 'Computers on the reef' (*Nature* **537,** 123–124; 2016) omitted to give the name of the system developed by Arjun Chennu and wrongly said that it is based on a neural-network algorithm. It is called HyperDiver, and it uses a machine-learning algorithm similar to the one used by CoralNet.

CAREERS

GRANTS Improve chances of success with a succinct case p.131

DIVERSITY Programme to support under-represented groups p.131

MENTORS An online networking tool fosters connections p.131

ADAPTED FROM CHUWY/GETTY

JOB APPLICATIONS

Under the covers

Having an impressive CV is one thing, but a well-written covering letter can really make you stand out from the crowd.

BY LESLEY EVANS OGDEN

Job-application letters (or cover or covering letters, as they're variously termed) may seem like mere formalities. But if you plan to apply for a science-related position, particularly in the academic, non-profit and industrial sectors, you need to write a spectacular one.

The document remains your first and best opportunity to act as both agent and salesperson for yourself: if it's done properly, only this component of your entire application package can simultaneously act as introduction, first-stage filter and cogent, compelling argument for your candidacy. Not until the interview — if you get one — will you have another chance to show why you are the best choice for the job.

Researcher applicants who want their cover letter to sparkle need to craft a document that's customized to the position. The letter should concisely explain how your competencies fit the criteria specific to the job, convey your excitement about the position and reveal some of your personality. It should also avoid hyperbole, typographical and other errors and exact duplication of points on your CV or résumé.

Some employers — particularly government agencies and organizations with a specialized online-only application process — do not welcome or use cover letters. But aside from these exceptions, it's best, hirers say, to include a letter, unless a job advert specifically bans it. The document remains an integral part of the recruitment process in industry and academia and for many non-profit organizations.

Why is it so important? Without one, say hiring managers, it can be tricky to identify the best candidates through their CVs and other application materials alone. These often start to sound drearily similar, says Karen Noble, head of research training and fellowships at Cancer Research UK (CRUK) in London, who frequently reviews applications for jobs in grant management or research administration. "Most people have done a PhD, they may have done a postdoc and they are now looking to be involved in administration," she says. "I want to see why they want to join the organization, and why this job." She says that an applicant's cover letter to CRUK should make it clear that the candidate has carefully studied the job ▶

description, and it should provide specific examples of how their skill set and experience meet the position's requirements.

NO ROBOTS

Hirers also stress that it is crucial to convey in your application letter that you've learnt as much as possible about the organization and specific job for which you're applying, and that you're not sending a generic submission. What a hiring manager at one company or organization may find powerful and persuasive in a cover letter may be viewed by their counterpart at another organization as irrelevant.

It's also important to spell out exactly how your abilities and interests align with the position, says Aaron Genest, who recruits candidates as part of his job with software firm Solido Design Automation in Saskatchewan, Canada. Because it's rare that an applicant's background exactly matches all of the criteria that hirers seek, it can help to make it clear in your letter that you are willing to do what is necessary to learn the specific skills that the hiring organization needs, such as by taking a course.

When all else is equal — background, education, skills, talents and abilities — getting the job is often down to your personality and how well you might fit in with the team. Outside an in-person interview, only your cover letter can offer a glimpse of your persona and disposition. Kevin Wang, a recruiter at biotech firm Stemcell Technologies in Vancouver, Canada, says that an applicant's 'personal brand', or individuality, is best conveyed in story form in the letter. You might, he suggests, write briefly about a time when you demonstrated your excellence at teamwork or problem solving, or explain in a concise way why you want the job. If you can link a personal interest to the position in some way, you should do so. Wang, who takes part in triathlons, says that if he were writing a cover letter for himself, he would probably include how triathlon training has taught him to be resilient and tenacious in the face of challenges.

Similarly, if you're enthusiastic and excited about the potential job, you should judiciously express that emotion. Cover letters often say things such as "I look forward to working with X", but you could express this more enthusiastically and with a bit more animation, says Iain Stenhouse, senior science director at the Biodiversity Research Institute in Portland, Maine. Cover letters that are vibrant and creative (but not outlandish) spur him to spend more time on the applicant's CV, he says. "They're where a candidate can really separate themselves from the pack."

INVESTIGATE OPTIONS

The importance of a cover letter may vary depending on whether you're applying for a position in industry, a non-profit organization or academia. So before agonizing over your letter, check to make sure it is needed at all. For example, cover letters are not part of the standard application package for some US federal government jobs, such as those at the US National Oceanic and Atmospheric Administration (NOAA). "I haven't seen a cover letter in years," says Richard Merrick, chief science adviser for NOAA Fisheries in Silver Spring, Maryland. Similarly, they are not used in the highly specialized hiring process at Diamond Light Source, the United Kingdom's synchrotron science facility in Didcot. Diamond's chief executive, Andrew Harrison, explains that the organization aims to standardize the hiring process, because some candidates who work with headhunters may not write the letters themselves.

Some academic institutions also do not consider cover letters to be crucial. Yvonne Buckley, hiring lead for zoology at Trinity College Dublin, says that it is only a single component of an application package, along with a CV and teaching and research statements. But although hiring committees may not read a letter if the other materials provide all of the necessary information, she says, candidates should not necessarily abandon the practice of including one. Especially in academia, where CVs can run to many pages, a cover letter can help to highlight achievements that relate to the job description and point committee members to where they can find more specific or detailed information.

NEGATIVE ATTENTION

It is important to remember that there is no line-by-line blueprint for a successful application document, save the need to tailor it to the hiring organization and the specific position. And although standing out is desirable, you do not want to do so for negative reasons. Recruiters and hiring managers warn that you need to make sure not to kick yourself out of the running because of mistakes or missteps that you could easily have avoided (see 'Tips for effective cover letters'), such as addressing the letter to the wrong person, making typographical or grammatical errors or including inadvertent leftovers from a previous application. "If someone is unable to express him- or herself without errors, that is an immediate reject," says Genest.

> ***"If someone is unable to express him- or herself without errors, that is an immediate reject."***

Another common issue is length. Outside academic environments, in which a two-page letter is common, recruiters emphasize that a carefully crafted one-page cover letter is enough. "A cover letter is not a book," says Monika Lips-Sandmeier, a human-resources specialist at the Swiss Federal Institute for Forest, Snow and Landscape Research in Birmensdorf.

A catalogue of your accomplishments, or anything else, will also act as a black mark against you. "Lists are deadly," says Genest. And although no one wants their application to be ignored, hiring managers warn that unorthodox attempts to stand out can backfire. At a careers fair, recruiter Lisa Knutson-Sealey once received a cover letter that was printed on fluorescent pink paper in bold type and a hard-to-read font. So, too, was the rest of the application. "It was just painful to look at," says Knutson-Sealey, who hires researchers and others for the Washington State Department of Ecology. It shouldn't really be a surprise to learn that the candidate did not get an interview. ■

Lesley Evans Ogden *is a freelance writer in Vancouver, Canada.*

IMPRESS EMPLOYERS

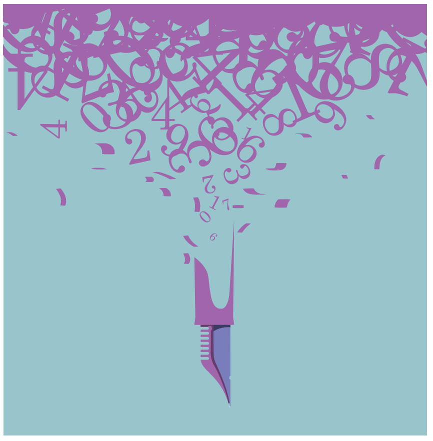
Tips for effective cover letters

- Address it to the appropriate person and, if necessary, do the homework to find out who that person is.
- Each cover letter needs to be carefully crafted for a specific job. Make sure to delete inadvertent leftovers from past application packages.
- Tell a story about why you are right for the position. Describe a previous job in which you used your problem-solving skills or demonstrated your ability to work as part of a team, for example.
- Convey your excitement and enthusiasm.
- Be honest and truthful. Don't exaggerate.
- Emphasize what doesn't get covered or rise to the surface in your CV or résumé. Expand on what makes you especially suitable, interesting or appealing for the specific position you are applying for.
- Proofreading for content, accuracy and style is key. Spell check and get a colleague or trusted personal contact to check spelling and readability, too. Automated spell checkers may not catch wrong words or homonyms such as pair/pare/pear. Be particularly careful about spelling the recruiter's or recipient's name correctly. Check any dates and addresses you are referencing.
- Avoid lists or bullet points.
- Be concise, and stick to a maximum of one page outside academia. For academic posts, two pages may be more acceptable.
- Avoid weird or unreadable fonts. L.E.O.

COLUMN

A better letter

When space is limited, make every word count, advises **Ingrid Eisenstadter**.



HONG LI/GETTY

Three years ago, the US-based grant-giving foundation that I work for decided to switch from asking for full grant proposals to asking instead for LOIs, variously called 'letters of inquiry' or 'letters of intent'. These are brief summaries of grant proposals. We made this decision mainly because an LOI is less time-consuming for applicants, an important consideration given that we are a small foundation that must turn away most of the proposals we receive — and also because it is less time-consuming for us.

Many of the LOIs we now receive have similar frailties: some give too much space to introductory discussions and fail to provide enough information about the research protocol. Some use too much technical vocabulary; others neglect to mention how much support they are looking for.

If the foundation does not ask you to specify the sum you are seeking, make sure you have done the research to know that you are within their funding range (and whether they fund internationally, if you live in a different country). If you are way above the giving limit, that is probably sufficient reason for them to turn away your enquiry.

As to technical vocabulary, do not assume that the reviewers will be conversant in the language of dozens of areas of speciality. In our last round of enquiries, we encountered PCL oil-soluble layers, Ancova analysis, fugitive dyes and elastic microbial repertoires, among other topics. Applications that clearly explained the terms began the full evaluation process at the outset, but those that were not clear had to wait for a time when a reviewer could research the terms or search for referees.

A large majority of grant seekers submit their LOIs immediately before the deadline, so you should keep in mind that those submissions will land in a pile-up as the calendar speeds towards the foundation's next board meeting. Thus, if your LOI gets a delayed reading, it risks not getting the detailed attention it deserves.

Most of the grant-makers that ask for LOIs have page-count or word-count limits. Our limit is 1,000 words, but I have seen some as low as 250 words. As you struggle to get your project summary down to that length, take some comfort in knowing that all applicants face the same problem, and sally forth.

LOI instructions do not usually address whether illustrations are welcome, although some researchers (not many) do include them. If your LOI would be much clarified by charts, graphs or photos, it is probably worth the gamble to include them. (If you must use an online application form, these uploads will probably not be accommodated.)

E-mail attachments can also pose a problem. When applicants send LOIs as attachments rather than in the body of the e-mail, the documents are often completely anonymous: they do not contain any identifying information. Many of the covering e-mails do not refer to the project title, and if the LOI is filed separately from the e-mail, it can create a serious problem.

So don't do that. Instead of just saying 'LOI attached' in your e-mail, provide a couple of sentences about what makes your proposal attractive or urgent, and include the title of your project and your full identification, including your e-mail address, in the LOI itself as well as in the body of the e-mail.

If you are invited to submit a full grant proposal after your LOI has been reviewed, you have cleared an important hurdle, and your chances of funding are better than when you submit a full grant proposal at the outset of the evaluation process. Leave yourself enough time to pass the final draft of your LOI around to colleagues and get their comments — and don't wait until the evening of the deadline day before you hit 'send'. ■

Ingrid Eisenstadter *is director of grants for the Eppley Foundation for Research in New York.*

GRANT AWARDS

Diversity boost

A recruitment and retention programme launched by the Howard Hughes Medical Institution (HHMI) in Chevy Chase, Maryland, aims to reduce barriers for women and under-represented minorities who seek academic-research careers in the life sciences. Those barriers include financial hardship, few mentoring and networking opportunities and uncertain career prospects. The Hanna H. Gray Fellows Program will support postdocs for up to four years with a US$60,000 annual stipend, plus up to $20,000 per year in flexible funds that can be used for family support or other purposes. For eligible fellows who become faculty members, the programme also provides up to $250,000 a year, plus $20,000 annually in flexible funds for up to four years. That funding will make recipients more attractive to universities, says Barbara Graves, HHMI senior scientific officer. Fellows can tap into HHMI's network of investigators for mentoring and networking, she says. The application deadline is 15 February 2017. See go.nature.com/2d9avh9 for more information.

NETWORKING

Match.com for mentors

The US National Institutes of Health (NIH) is rolling out an online tool to match early-career researchers with mentors. The online platform, MyNRMN, asks would-be mentees and mentors to create profiles and connect through direct messages or by joining groups that share their interests. Postdocs, faculty members, staff researchers and administrators can be mentors; mentees can be students and postdocs. The tool is part of the NIH-funded National Research Mentoring Network, which was created in 2014 to improve networking and mentorship opportunities for scientific trainees from diverse backgrounds. More than 2,000 mentees and 1,000 mentors have registered with the programme, says Jamboor K. Vishwanatha, a molecular biologist at the University of North Texas Health Science Center in Fort Worth who is developing the network. Mentees can connect with multiple mentors who share their research interests, location or ethnicity. The platform can also be used to assess the interactions that correlate with later success in scientific careers. "This will allow the development of a 'science' of mentoring and help us to develop best practices," says Vishwanatha.

SIMPLE THINGS

A question of survival.

BY REBECCA BIRCH

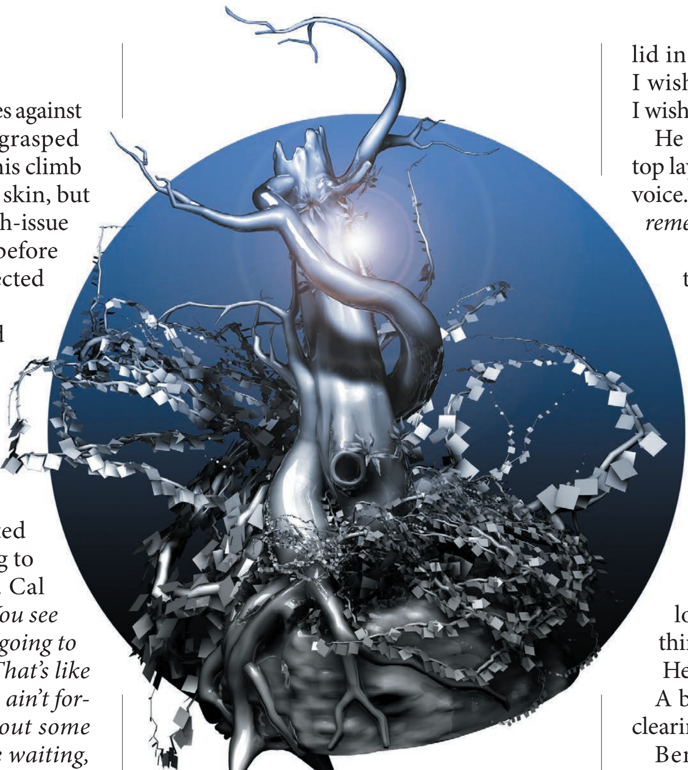


ILLUSTRATION BY JACEY

Ben Nabors curled his bare toes against the rough stone face and grasped the top of the cliff. Once, this climb would have torn shreds from his skin, but that was long ago — when his Earth-issue boots finally wore through and before he built up the calluses that protected him now.

The satchel, stitched from old shirts, shifted against his spine. "Almost there, Cal."

Ben levered himself over the edge and paused to catch his breath.

Old Lookout stood at the far end of the precipice, its twisted trunk and stunted limbs clinging to the rock, stubborn and defiant. Cal had taken the tree as a mascot. *You see Old Lookout, Ben? Ain't nothing going to move it or call it a blamed fool. That's like you and me. The folks back home ain't forgotten us. They just gotta work out some kinks, that's all. And us? We'll be waiting, just like Old Lookout, and I'll get home to my Abigail.*

Ben hadn't had the heart to tell Cal he'd given up long ago. After the last of the projected emergency portal dates had passed with no sign of contact — after Mindy, with her dirt-painted hands and soft blue eyes, had succumbed to the summer fever — Ben knew they weren't going home. Not tomorrow. Not the next day. Not ever.

The cookpot filled with Cal's ashes dragged at Ben, as if, instead of his friend's remains, it held some tiny singularity, pulling the world into its gravitational well. Ben's throat clogged and he blinked away tears. They wouldn't do any good. Besides, he'd given up on caring after Mindy died. It was easier that way. Eat, drink, sleep. Try not to get killed. Simple things.

At least, that's what he told himself.

Cal might've been an irritatingly optimistic bugger, but he was smart. *You ain't fooling me with your 'I don't care 'bout nothing'. Coulda taken off to try to find some better place, but you're still here. You're just as much Old Lookout as me.*

Ben cleared his throat and focused on his feet. One step after another until he reached Old Lookout's base. He touched the trunk with one hand and edged towards the precipice. The bark against his palm felt warm. Alive.

Taking a cautious step back, Ben took off the satchel and pulled out the cookpot. His warped reflection stared back at him, all unkempt hair and beard, with eyes so blank they might've already been dead.

When Cal died, Ben knew exactly where he'd spread the ashes, though he'd only climbed the precipice once before. A wild, terrifying rush, his fingers clinging in small clefts, booted feet scrabbling over barely glimpsed protrusions. An escape from the others, when the second planned portal event had passed with no sign of any attempt from Earth. A challenge, to prove he was still worth something. An opportunity to die without shame, if he should fall.

But he hadn't fallen. Cal had seen him from far below and waved up at him with both arms, jumping and hollering …

Dwelling on the memory wouldn't do anybody any good now. Cal was gone, the last thing still anchoring Ben to this world.

The pull of the singularity contained in the cookpot dragged away what little colour and substance remained that was still Ben Nabors, leaving him empty and weak. He dropped to his knees.

Ben untied the bootlace holding the lid in place. "You were a good man, Cal. I wish there were a way to tell someone. I wish …"

He lifted the lid. A crisp breeze raked the top layer of ash. He could almost hear Cal's voice. *All you gotta do is wait. Old Lookout, remember?*

Ben leaned forward and glanced over the edge. The fall didn't frighten him. At the bottom there would be an *end*.

His fingers tightened on the cookpot's rim. "I'm sorry, Cal. You were wrong about me."

He rose to his feet and tipped the pot towards Old Lookout's base. Cal's ashes rained down over the clinging roots and gathered into drifts like snow.

Ben turned away and looked out over the abyss. One little step and his long exile would be over. One simple thing.

He lifted his foot.

A bright light glimmered down in the clearing by the lake.

Ben blinked and rubbed his eyes — surely he was dreaming — but when he opened them again, the light was still there, intensifying, the telltale periwinkle hue unmistakable.

A sharp breeze blew, throwing off his balance.

He grabbed instinctively for Old Lookout. The same warmth he'd felt earlier swept through his fingers, his arm and deep into his heart. All the vitality and hope that had vanished into the singularity slammed back into him so hard his whole body trembled.

Protocol said a portal should be open for two hours, but protocol had failed him for more years than he could count. What if the portal failed before he could reach it? What if it was nothing more than a hallucination conjured by a mind left to die alone? Was it worth the risk of caring enough to try?

A gritty breeze brushed against his face — the last remnants of Cal, who'd left his little girl behind. Ben straightened and drew in a breath. He could do this. He *had* to.

Heart racing, Ben lowered himself over the cliff edge. Foothold after foothold. Hand over hand. For Cal and his Abigail.

Simple things. ■

Rebecca Birch *is a Seattle resident who has been published in markets including* Nature, Cricket *and* Flash Fiction Online. *You can find her online at www.wordsofbirch.com.*